Advances in solar system magnetohydrodynamics

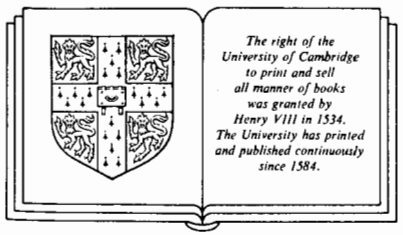
Edited by

Eric R. PRIEST

and

Alan W. WOOD

Department of Mathematical and Computational Sciences
University of St Andrews



CAMBRIDGE UNIVERSITY PRESS

Cambridge

New York *Port Chester*

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10022-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1991

Printed in Great Britain at the University Press, Cambridge

British Library cataloguing in publication data available

Library of Congress cataloguing in publication data available

ISBN 0 521 40325 1 hardback

Contents

2

Dynamo models and Taylor's constraint

C. A. JONES

University of Newcastle-upon-Tyne, NE1 7RU, U.K.

2.1 Introduction

Many of the basic concepts of magnetic field generation in conducting fluids are now understood, but the construction of mathematical solutions to the MHD equations which model planetary and stellar dynamos remains a challenging task. Even if it is assumed that the dynamo is powered by a simple form of buoyancy-driven convection, to find a plausible dynamo model we need to solve the equation of motion for the fluid velocities and the induction equation for the magnetic fields in a fully three-dimensional configuration. If this is achieved, we then face the further difficulty that the parameters appropriate for planetary and stellar dynamos are not generally those which are easiest to handle numerically. Because of these difficulties, a step by step approach is required: we start with models in simplified geometry which examine one aspect of the problem, and, as experience is gained, we hope to develop models in realistic geometries which combine many different aspects. So, for example, most of the models described here are based on thermal convection: our understanding is not yet sufficiently developed for it to be worthwhile to incorporate compositional convection effects (Fearn, 1989), which are less well understood.

There are, naturally, a number of different routes towards the goal of developing realistic dynamo models, and in this chapter we review only models based on the rapid rotation limit. In this approach it is assumed that the inertial terms in the equation of motion are negligible, and that viscous forces are small. As we shall see, neglecting viscous forces everywhere can lead to difficulties, but the philosophy we adopt is to neglect them wherever possible. A different approach to the MHD dynamo problem has been pioneered by

Busse (1975, 1976 and his chapter in this volume) and Zhang and Busse (1988, 1989, 1990), in which the rapid rotation limit is not taken at the start, but is approached gradually by increasing the Taylor and Rayleigh numbers in the full equations. Ultimately, the connections between these two different approaches should become clear, but this is still some way off. It is not yet possible to get the spatial resolution necessary for high Taylor and Rayleigh number solutions of the full equations, although rapid advances in computing technology and algorithm development hold out some hope for the future. The problems in which the rapid rotation limit is perhaps most natural are those of the Earth's dynamo and dynamos in the major planets. In the solar dynamo, the Rossby number $U/L\Omega$ turns out to be of order unity if the conventional mixing length estimates of convective velocities are assumed. The application of the rapid rotation limit is therefore not so clear in the solar case: nevertheless, the actual magnitude of the turbulent velocity fluctuations in the lower solar convection zone is not known with any certainty, and it may be that even if the Rossby number is not very small, the rapid rotation limit may give some insight into the large scale dynamics of the solar convection zone (Jones, Roberts and Galloway, 1990). Summaries of the geophysical applications of the ideas presented in this review have been given by Gubbins and Roberts (1988), Roberts and Gubbins (1988), and Roberts (1988a). Progress towards understanding the nonlinear dynamics of the rapid rotation limit has mainly been on two fronts: the magnetoconvection problem in which an initial magnetic field is imposed and the form of the convection is analysed, and the axisymmetric α-effect dynamo problem. The organisation of this chapter is divided along these lines, although it will be seen that there are many similarities in the two problems. A successful synthesis of these two problems has not yet been satisfactorily achieved, but progress has been made in uncovering some of the difficulties such a synthesis encounters (Fearn and Proctor, 1987). A comprehensive review of the basis of much of the discussion here has been given by Fearn, Roberts and Soward (1988); other reviews covering similar ground to this one are by Soward (1989) and Roberts (1988b). The fundamental processes of dynamo action are discussed in the books of Moffatt (1978) and Parker (1989). Parker's book also contains much information about the application of these ideas to stars and planets.

2.2 The rapid rotation limit

If we consider Boussinesq fluid, convecting in the presence of rotation and magnetic field, the equation of motion can be written as

$$Ro[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u}] + \hat{\mathbf{z}}\times\mathbf{u} = -\nabla p + \Lambda \mathbf{j}\times\mathbf{B} + R\theta\hat{\mathbf{r}} + E\nabla^2\mathbf{u} \quad (2.1)$$

the magnetic induction equation as

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla\times(\mathbf{u}\times\mathbf{B}) + \nabla^2\mathbf{B} \quad (2.2)$$

and the temperature equation as

$$\frac{1}{q}[\frac{\partial \theta}{\partial t} + (\mathbf{u}\cdot\nabla)\theta] = (\mathbf{u}\cdot\hat{\mathbf{r}}) + \nabla^2\theta \quad (2.3)$$

We also have

$$\nabla\cdot\mathbf{B} = 0, \qquad \nabla\cdot\mathbf{u} = 0, \qquad \mathbf{j} = \nabla\times\mathbf{B}/\mu. \quad (2.4)$$

These equations are all based on dimensionless variables, taking as unit of length d, which is usually the core-radius in the geodynamo problem, or the layer depth in plane layer models. The unit of time is d^2/η, where η is the magnetic diffusivity. For the geodynamo, this unit of time is around 10^5 years. The velocity is then scaled on $\mathcal{U} = \eta/d$, the magnetic field on some value B_0 ,and the temperature on βd, where β is the temperature gradient in the conducting state. $\hat{\mathbf{r}}$ and $\hat{\mathbf{z}}$ are unit vectors in the radial direction and rotation direction respectively. The dimensionless parameters occurring in these equations are the Rossby number, Ro, the Elsasser number, Λ, the Rayleigh number, R, the Ekman number E and the Roberts number q, where

$$Ro = \frac{\mathcal{U}}{2\Omega d}, \qquad \Lambda = \frac{B_0^2}{2\mu\rho\Omega\eta}, \qquad R = \frac{g\alpha\beta d^2}{2\Omega\kappa},$$

$$E = \frac{\nu}{2\Omega d^2}, \qquad q = \frac{\kappa}{\eta}.$$

The density is ρ, α is the expansion coefficient, g is the acceleration due to gravity, Ω the rotation rate and κ, ν are the thermal and viscous diffusivities, respectively. We have assumed all diffusivities are constant; μ is the permeability. The rapid rotation limit is obtained by letting $Ro \to 0$, and $E \to 0$, while R and Λ remain of order unity. As mentioned above, it is not always possible to set $E = 0$, but we always consider $Ro = 0$ and E small.

When the rapid rotation approximation is not made, the parameters commonly used to describe the convection are the Rayleigh number, Ra, the Taylor number Ta, and the Chandrasekhar number

Q, given by

$$Ra = \frac{g\alpha\beta d^4}{\kappa\nu}, \quad Ta = \frac{4\Omega^2 d^4}{\nu^2} = E^{-2}, \quad Q = \frac{B_0^2 d^2}{\mu\rho\eta\nu}.$$

Note that this Rayleigh number Ra is different from the one we defined earlier, R. In the rapid rotation limit all these quantities go to infinity (we are thinking of the diffusivity ratios as fixed in this limit) in such a way that

$$Ra \sim Ta^{\frac{1}{2}} \sim Q.$$

The effect of taking the rapid rotation limit is to discard the inertial term in (2.1) and to ignore the viscous term wherever possible. If we set $E = 0$ we obtain the magnetostrophic balance equation

$$\hat{\mathbf{z}}\times\mathbf{u} = -\nabla p + \Lambda \mathbf{j}\times\mathbf{B} + R\theta\hat{\mathbf{r}}. \tag{2.5}$$

This equation no longer uniquely determines $\mathbf{u}$. To any solution we can add a geostrophic velocity $U_G(s)$, s being the distance from the rotation axis; the Coriolis effect of this velocity can be balanced by a pressure gradient. One way of removing this degeneracy is to retain some coupling between the core and the mantle. The simplest assumption is that this coupling is dominated by the contribution from the Ekman boundary layer which leads to (Tough and Roberts, 1968)

$$E^{\frac{1}{2}}\Lambda^{-1}\left(\frac{2}{z_B}\right)^{\frac{1}{2}} \pi s U_G = \int_S \mathbf{j}\times\mathbf{B}\cdot\hat{\boldsymbol{\phi}}\, ds. \tag{2.6}$$

where S is a cylinder of radius s which cuts the sphere of radius unity at $z = \pm z_B$. A more complicated, but probably more realistic equation results if the coupling is provided by an electromagnetic torque produced by a finite mantle conductivity (Braginsky, 1975; Cupal and Hejda, 1989). An important consequence of replacing (2.1) by (2.5) and (2.6) is that Rossby waves, inertial and Alfvén waves, which have periods short compared to the magnetic diffusion times are filtered out. Omitting such modes may seem a serious drawback to the rapid rotation approximation; however, since the dynamo processes in the core act over a timescale of thousands of years it will be an immense task to resolve oscillations with periods less than a year. The complexity of the MHD equations means that in the foreseeable future computing resources are going to be severely stretched finding any realistic solutions: in view of this, it seems sensible to filter out fast modes at the outset. Alfvén waves in the earth's core have a longer time-scale than Rossby or inertial waves; torsional oscillations (standing Alfvén waves) have periods in the range 10-100 years, still

short compared to the magnetic diffusion time. They are usually neglected, but can be restored by incorporating a term containing the time-derivative of U_G on the left-hand side of (2.6); see the article by Fearn, Roberts and Soward (1988) for further details.

2.3 Magnetoconvection

2.3.1 Linear Theory

The behaviour of convection in the rapid rotation approximation with a given magnetic field has been studied in a variety of different geometries. The linear Boussinesq problem in Cartesian geometry was looked at by Eltayeb (1972, 1975); a slab is considered between planes at $z = 0, d$. Gravity is in the $-z$ direction and the imposed magnetic field is in the x direction. The rotation vector can lie anywhere in the $y - z$ plane, but the two cases studied most are (a) $\boldsymbol{\Omega} = \Omega_0\hat{\mathbf{z}}$, which mimics the behaviour near the polar regions and (b) $\boldsymbol{\Omega} = \Omega_0\hat{\mathbf{y}}$ which relates more to conditions near the equator. The compressible problem with $\boldsymbol{\Omega} = \Omega_0\hat{\mathbf{y}}$ has been examined by Jones, Roberts and Galloway (1990). The linear Bousinesq problem in a sphere was analysed by Eltayeb and Kumar (1977), Fearn (1979a,b) and Fearn and Proctor (1983). The most popular choice of magnetic field in these studies is to have a toroidal field $\mathbf{B} = B\hat{\boldsymbol{\phi}}$, with B proportional to s. This corresponds to a uniform current in the z direction, and hence satisfies $\nabla^2\mathbf{B} = 0$. If dynamo action is taking place, the generated field will of course be more complicated than this, and so the possibility of field gradient instabilities arises (Acheson, 1983, Fearn and Weigelhofer 1990).

There are some general features which are common to all these models, and some which are strongly model-dependent. In all models, a key parameter is the Elsasser number, Λ, which measures the relative strength of Lorentz and Coriolis forces. If this number is small, rotation constrains the convection rolls so that their axes are as near parallel to the rotation vector as the geometry permits; we would expect this, since this situation approximates to that in which the Proudman-Taylor theorem applies. As the Elsasser number is raised, the direction of the axes of the preferred rolls changes, the wavelength of the rolls increases, and the critical Rayleigh number falls. Convection generally occurs at minimum Rayleigh number when the Elsasser number is of order unity. If the value of Λ is further increased, the magnetic field starts to constrain the convection

strongly, and the critical Rayleigh number begins to increase with field strength.

Boussinesq plane-layer convection normally onsets at a pitchfork bifurcation (exchange of stabilities) provided the diffusivity ratio $q = \kappa/\eta$ is not much greater than unity: if it is, a Hopf bifurcation may occur first (Eltayeb 1975). In spherical geometry, or when compressibility is present, convection in the rapid rotation limit always occurs via a Hopf bifurcation. The question of wave speed, and even of the direction of wave propagation, appears to be very model-dependent.

2.3.2 Nonlinear Theory and Taylor's Constraint

To date, nonlinear investigations of magnetoconvection in the rapid rotation limit have been in simple geometry: infinite plane layers (Roberts and Stewartson, 1974, 1975), ducts (Soward, 1986; Jones and Roberts 1990) or cylinders (Skinner and Soward, 1988).

If viscosity is ignored entirely, nonlinear models run into difficulty immediately, and only in exceptional cases can weakly nonlinear solutions be found in the neighbourhood of the linear solutions. The origin of the difficulty lies in the magnetostrophic balance equation (2.5). If the linear solution for the perturbed field is $\mathbf{b}'$, and $\mathbf{B} = B_0\hat{\boldsymbol{\phi}} + \mathbf{b}'$ is returned into (2.5) it proves to be impossible to balance all the quadratic interaction terms generated by the Lorentz force. This is connected with the fact that the geostrophic velocity is not determined by the magnetostrophic equation; we again take an integral of the azimuthal component of the momentum equation (2.5). In spherical and cylindrical geometry this integral is over cylinders whose axis is parallel to the rotation axis, while in duct geometries the integral is over planes which contain the rotation axis and the magnetic field vector. When this integral is taken, the contribution from the Coriolis term vanishes because there can be no net flow across the cylinders (or planes), and the pressure and buoyancy forces do not contribute either. Therefore, if magnetostrophic balance holds, Taylor's condition,

$$\int_s \mathbf{j}\times\mathbf{B}\cdot\hat{\boldsymbol{\phi}}\,ds = 0, \tag{2.7}$$

or its equivalent in Cartesian geometry, must hold. It was pointed out by Roberts and Stewartson (1975) that the quadratic interactions arising from the linear solutions do not in general satisfy (2.7). In a normal weakly nonlinear expansion, quadratic interaction terms

are balanced by corrections to the velocity, temperature, etc. at the $O(\epsilon^2)$ level. In the magnetostrophic problem this is usually impossible, as only the magnetic field comes into the Taylor constraint and the quadratic interaction of the linear terms will not balance the second-order perturbation of the magnetic field. It should be noted, however, that the problem of a single roll in infinite plane layer geometry with the rotation axis vertical is exceptional (Roberts and Stewartson 1974), in that its linear solution happens to satisfy Taylor's constraint exactly.

To resolve this difficulty we must add extra terms to the equation of motion, as mentioned in Section 2. In the geodynamo problem, the most natural term to balance this ϕ-component of the Lorentz force is a core-mantle coupling term, either electromagnetic, topographic or viscous. In problems such as stellar dynamo, where there are no obvious rigid boundaries, we can restore a turbulent viscosity throughout the dynamo region to provide the balance. Viscous coupling through an Ekman boundary layer provides the simplest model, so that although electromagnetic or topographic coupling is probably more realistic, viscous coupling is most often used in models.

If we further assume that this core-mantle coupling is weak, then only a weak Lorentz force can be allowed, so that the variation of the magnetic field from its initial state value has to be small. If we introduced a normalised Ekman number

$$\epsilon = E\Lambda^{-2}2\pi^2$$

(2.6) can be written

$$\epsilon^{\frac{1}{2}}U_G = \frac{(1-s^2)^{\frac{1}{4}}}{s}\int_s \mathbf{j}\times\mathbf{B}\cdot\hat{\boldsymbol{\phi}}\,ds. \tag{2.8}$$

The geostrophic wind $U_G(s)$ is $O(1)$, so that (2.8) implies that the perturbation magnetic field $\mathbf{b}'$ and velocities satisfy

$$\frac{\mathbf{b}'}{B_0} = O(\epsilon^{\frac{1}{4}}), \qquad \mathbf{u}' = O(\epsilon^{\frac{1}{4}}).$$

This *weak coupling* approximation can be viewed as part of the rapid rotation approximation if the coupling is viscous, as the coupling parameter is dependent on the Ekman number. If other forms of coupling are involved, however, this link is broken, and it is consistent to have strong coupling, with $\epsilon = O(1)$, even with rapid rotation. The actual value of ϵ in the geodynamo is not entirely clear. Estimates of the order of magnitude of electromagnetic coupling are dependent on lower mantle conductivity, about which little is known: estimates are

generally rather low (Gubbins and Roberts, 1988). Topographic coupling (Hide, 1966) is potentially much larger, but to date there is no direct evidence for it. However, if the presently unexplained decade variations in the length of the day are due to variation of angular momentum in the core, this suggests a "strong" value of core-mantle coupling. The weak coupling equations can now be written

$$\mathbf{B} = B_0\hat{\boldsymbol{\phi}} + \epsilon^{\frac{1}{4}}\mathbf{b}', \quad \mathbf{u} = U_G\hat{\boldsymbol{\phi}} + \epsilon^{\frac{1}{4}}\mathbf{u}', \quad \theta = \epsilon^{\frac{1}{4}}\theta', \tag{2.9}$$

where $\mathbf{b}'$ and $\mathbf{u}'$ satisfy

$$\boldsymbol{\nabla}\cdot\mathbf{b}' = \boldsymbol{\nabla}\cdot\mathbf{u}' = 0, \tag{2.10}$$

and the induction equation becomes

$$\frac{\partial \mathbf{b}'}{\partial t} = \boldsymbol{\nabla}\times(U_G\hat{\boldsymbol{\phi}}\times\mathbf{b}') + \boldsymbol{\nabla}\times(\mathbf{u}'\times\mathbf{B}_0) + \boldsymbol{\nabla}^2\mathbf{b}', \tag{2.11}$$

the equation of motion divides into two parts,

$$\hat{\mathbf{z}}\times\mathbf{u}' = -\boldsymbol{\nabla}p' + \Lambda\{(\boldsymbol{\nabla}\times\mathbf{b}')\times\mathbf{B}_0 + (\boldsymbol{\nabla}\times\mathbf{B}_0)\times\mathbf{b}'\} - R\theta'\hat{\mathbf{r}} \tag{2.12}$$

and

$$U_G(s) = \frac{(1-s^2)^{\frac{1}{4}}}{s}\int_S (\boldsymbol{\nabla}\times\mathbf{b}')\times\mathbf{b}'\cdot\hat{\boldsymbol{\phi}}\,ds \tag{2.13}$$

and the temperature equation becomes

$$\frac{1}{q}\left[\frac{\partial\theta'}{\partial t} + (U_G\hat{\boldsymbol{\phi}}\cdot\boldsymbol{\nabla})\theta'\right] = (\mathbf{u}'\cdot\hat{\mathbf{r}}) + \nabla^2\theta'. \tag{2.14}$$

Note that equations (2.10-2.12) and (2.14) are linear in the dashed quantities; the only nonlinearity comes in through the geostrophic wind, determined by (2.13). These linear equations can be thought of as describing magnetoconvection in the presence of a shear, $U_G(s)$. Indeed, if $U_G(s)$ were prescribed, we would have a standard linear eigenvalue problem for the growth rate as a function of wavenumber. Generally, an imposed shear in the linear problem makes convection more difficult to achieve. If R_c denotes the critical Rayleigh number for the convection with no shear, then with an imposed shear we obtain marginal convection with $R > R_c$. In the nonlinear problem at given R the form of the shear is determined by (2.13), while its magnitude is determined by the requirement that the linear part of the problem be exactly marginal at R. Once the magnitude of the shear is set by this requirement, the magnitudes obtained by the convective quantities $\mathbf{u}'$, $\mathbf{b}'$ and θ' are set by equation (2.13). Because the amplitude of the convection is scaled on $\epsilon^{\frac{1}{4}}$, which is determined by the core-mantle coupling, we call the regime governed by (2.9) to (2.14) the *coupling regime*. This coupling regime persists unless a *Taylor state* is reached. The shear builds up as R is increased, and so

the form of the convective quantities, and in particular $\mathbf{b}'$, changes as R increases. In some problems, $\mathbf{b}'$ develops towards a configuration in which it satisfies Taylor's constraint at some Rayleigh number R_T. As R approaches R_T from below, $U_G(s)$ still stays non-zero and of $O(1)$; indeed, U_G does not vary particularly in the neighbourhood of R_T . What happens is that the magnitude of $\mathbf{b}'$ rapidly increases, so that although the spatial structure of the integrand of

$$\tau = \frac{(1-s^2)^{\frac{1}{4}}}{s}\int_S (\boldsymbol{\nabla}\times\mathbf{b}')\times\mathbf{b}'\cdot\hat{\boldsymbol{\phi}}\,ds \tag{2.15}$$

is such that the positive and negative contributions are almost exactly equal, τ remains nearly constant. The near cancellation of the integral as $R \to R_T$ is exactly compensated for by the big increase in the magnitude of $\mathbf{b}'$.

This large increase of $\mathbf{b}'$, and hence $\mathbf{u}'$, means that we can no longer ignore terms of the form $(\mathbf{u}'\cdot\boldsymbol{\nabla})\mathbf{b}'$ and $(\mathbf{u}'\cdot\boldsymbol{\nabla})\theta'$. Once we get to $R > R_T$, we enter a new asymptotic regime, and a new scaling is required. $\mathbf{b}', \mathbf{u}'$ and θ' are now $O(1)$ quantities, and the relevant equations are (2.2), (2.3), (2.4) and (2.5) together with (2.7). In this regime the geostrophic wind U_G can adjust the magnetic field $\mathbf{B}$ through the induction equation in such a way that Taylor's constraint (2.7) is exactly satisfied. For Rayleigh numbers $R > R_T$, we can now have solutions which are independent of the core-mantle coupling. This does not imply that core-mantle coupling is unimportant in the geodynamo problem; what we are saying here is that if we assume the coupling is weak, then for $R > R_T$ we have mathematical solutions which are to leading order independent of any core-mantle coupling. To examine the effect of any such weak coupling we could look at how this leading order solution is perturbed by the inclusion of small extra terms in (2.7). All this presupposes that a second critical Rayleigh number R_T exists; this is not obvious. A similar question occurs in α-effect dynamo problems, but in rapidly rotating magnetoconvection the only geometries studied so far are plane and cylindrical systems. The duct model with rotation and gravity parallel (see Figure 2.1a) was analysed by Soward (1986); the case with rotation and gravity perpendicular (Figure 2.1b) was looked at by Jones and Roberts (1990) and the cylindrical problem with rotation and gravity parallel (Figure 2.1c) by Skinner and Soward (1988). In the geometry of Figure 2.1a, the surfaces for the Taylor integral are the planes with x constant, while in Figure 2.1b they are the planes with z constant. In both studies with rotation and gravity parallel a second critical

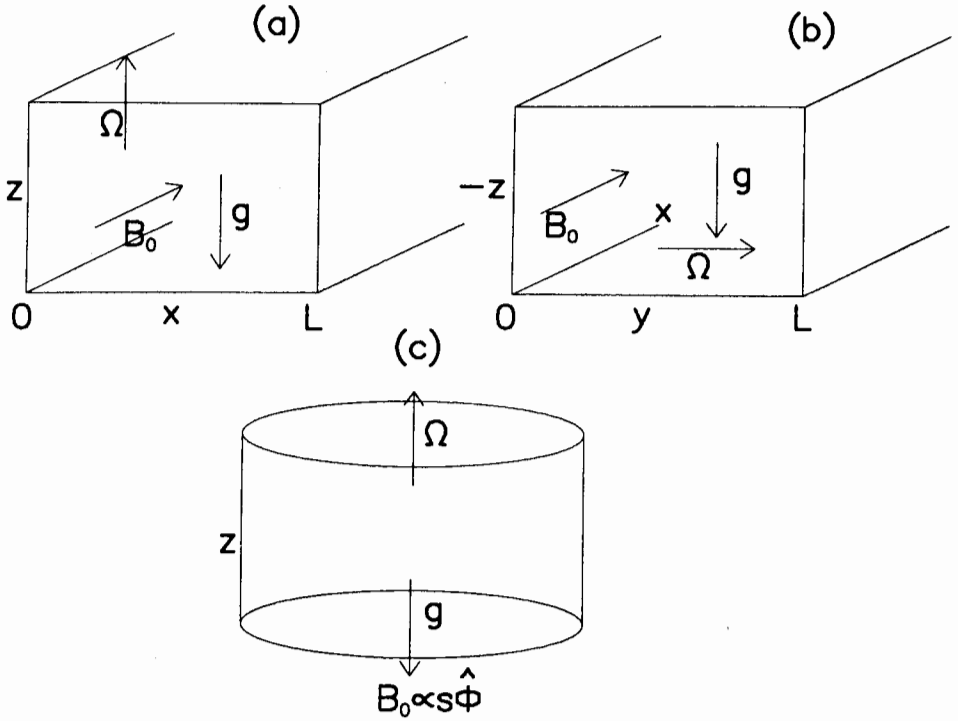


Figure 2.1. The geometry of duct and cylindrical models.

Rayleigh number R_T was found and the scenario described above is appropriate. However, the picture is not so simple at small $q = \kappa/\eta$, a limit which is appropriate for the geodynamo if no turbulent eddy heat transport is envisaged. In that case, the solutions at $R = R_T$ have a rather complicated spatial structure (Soward, 1986) dividing the duct into two regions: in one region the terms are scaled as in the $q = O(1)$ case with velocity scale of order η/d, but in the other region the geostrophic wind $U_G(s)$ is $O(q^{-1})$ and large, while the temperature fluctuation θ' is small (being reduced by the large shear) so that buoyancy is negligible there.

The behaviour of the duct model with $\mathbf{\Omega}$ and $\mathbf{g}$ perpendicular (Figure 2.1b) is strikingly different. Convection onsets with the solutions proportional to $\exp i(kx + ly)$, where $l = n\pi/L$, L being the duct-width. Provided the Elsasser number is larger than a value close to 4, minimum Rayleigh number occurs at some non-zero wavenumber

$k = k_c$, with $n = 1$. For Λ less than about 4, modes with $l = 0$ are preferred and the story is different: here we describe only the larger Λ behaviour. If the wavenumber k is kept constant at k_c as the Rayleigh number is increased (the conventional practice in convection theory) then it can be shown that no Taylor state can be reached whatever the Rayleigh number. This result also holds for a variety of magnetic boundary conditions; it does, however, depend on the wavenumber k not increasing much as R is increased. A possible escape route may come from an examination of the stability of the nonlinear solutions for varying k. It may be that there is a tendency for larger k solutions to be more stable as R increases: if this is the case, this duct model may also evolve towards a Taylor state by shortening the wavelength in the magnetic field direction.

The duct model shown in Figure 2.1b was also examined for the case when viscosity in the interior balances the Taylor integral. This occurs when stress-free boundary conditions are imposed rather than rigid-lid conditions. This case was suggested by the possibility of dynamo action in rapidly rotating stellar convection zones. The scalings in this case are slightly different, with $\mathbf{u}'$ and $\mathbf{b}'$ being $O(\epsilon^{\frac{1}{2}})$ instead of $O(\epsilon^{\frac{1}{4}})$, but otherwise the behaviour is quite similar. However, the stress-free case does undergo a symmetry-breaking bifurcation as R is increased: the steady convection loses stability to two travelling wave solutions going in opposite directions. This illustrates a phenomenon which can also occur in α-effect dynamos: when the second critical Rayleigh number, R , (or dynamo number D) lies well above R_c, there is a possibility that secondary bifurcations will occur before R reaches R_T; this may lead to additional complications. We cannot expect that the simple scenario of first transition at R_c , followed by a smooth development through the coupling regime ending at a Taylor state, will be universal in all rapidly rotating approximation problems.

Naturally, we would like to solve the weak coupling equations in a spherical geometry: this is, however, a substantially more difficult task than any of the problems that have been considered so far. Because of the particular form of the nonlinearity in the coupling regime, in all the geometries of Figure 2.1 the variables have a particularly simple spatial dependence: in Figure 2.1a the dependence is $\exp ily \sin\pi z$, in Figure 2.1b variables are proportional to $\exp(ikx + ily)$ and in Figure 2.1c the dependence is $\exp im\phi \sin\pi z$. All have solutions which are either time-independent or a simple sinusoidal oscillation: in consequence they all give rise to ordinary

differential equation eigenvalue problems which can generally be handled by standard numerical methods (although things are not quite straightforward at small q). The spherical geometry problem separates spatially into $f(r,\theta)\exp im\phi$ but no further. Also, additional numerical complications result from the Taylor surfaces being cylinders while the coordinate system is spherical. It is, however, important that these difficulties be overcome, as the magnetoconvection problem is a key ingredient in the process of producing self-consistent dynamos. Its importance lies in the fact that the nonaxisymmetric parts of the solution are the elements which generate poloidal field from toroidal field. In the absence of proper solutions to the magnetoconvection problem, this process has to be incorporated through an α-effect. The difficulty with this is that a more or less arbitrary choice has to be made for the magnitude and distribution of the α-term.

2.4 Axisymmetric dynamo models

2.4.1 α^2 and $\alpha\omega$ Dynamos

Cowling's well known antidynamo theorem (1957) means that we cannot construct purely axisymmetric dynamos, essentially because the poloidal field cannot be maintained. Nonaxisymmetric motions which produce poloidal field must therefore be incorporated into our axisymmetric framework by some assumptions. We can write

$$\mathbf{u} = \overline{\mathbf{U}} + \mathbf{u}', \qquad \mathbf{B} = \overline{\mathbf{B}} + \mathbf{b}',$$

where $\overline{\mathbf{U}}$ and $\overline{\mathbf{B}}$ are the azimuthal averages, and $\mathbf{u}'$ and $\mathbf{b}'$ are the asymmetric parts. When these expressions are put into the induction equation, azimuthal averaging gives

$$\frac{\partial \overline{\mathbf{B}}}{\partial t} = \nabla\times(\overline{\mathbf{U}}\times\overline{\mathbf{B}}) + \nabla\times\overline{(\mathbf{u}'\times\mathbf{b}')} + \nabla^2\overline{\mathbf{B}} \tag{2.16}$$

$\nabla\times\overline{(\mathbf{u}'\times\mathbf{b}')}$ is the average of the asymmetric terms in (2.16) which can sustain a poloidal field. If the non-axisymmetric terms are small compared to the axisymmetric term, and the mean field is primarily toroidal, Braginsky (1964) showed that

$$\overline{(\mathbf{u}'\times\mathbf{b}')} = \alpha B\hat{\boldsymbol{\phi}} \tag{2.17}$$

where α is an axisymmetric function of position related to the helicity of the flow (Soward, 1972).

An alternative procedure is to regard $\mathbf{u}'$ and $\mathbf{b}'$ as fluctuating quantities which vary over a short length scale, which is then averaged

over in a Mean Field approximation (e.g. Moffatt, 1978; Krause and Rädler, 1980). We then get

$$\overline{(\mathbf{u}'\times\mathbf{b}')} = \alpha\mathbf{B} \tag{2.18}$$

α again being an axisymmetric function of position. Equations (2.17) and (2.18) are actually very similar if magnetic field is predominantly toroidal, as is probably the case in the Earth and the Sun.

Dynamo models fall into two classes: α^2 models in which the job of generating both toroidal and poloidal field is done by the α-effect terms, and $\alpha\omega$ models where the toroidal field is generated by differential rotation stretching the poloidal field lines. One can, of course, have models in which both mechanisms contribute to toroidal field generation.

Kinematic dynamo theory consists of solving (2.16) with (2.18) so that the induction equation becomes

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla\times(\mathbf{u}\times\mathbf{B}) + \nabla\times\alpha\mathbf{B} + \nabla^2\mathbf{B} \tag{2.19}$$

$\mathbf{u}$ and α being prescribed functions of position. Many kinematic dynamo models have been developed, in many different geometries; a review of many of these is given in Parker's (1979) book.

Since the differential rotation is important in many dynamo models, we must examine the equation of motion. The ϕ-component of the curl of equation (2.5),

$$\frac{\partial u_\phi}{\partial z} = \Lambda\hat{\boldsymbol{\phi}}\cdot\nabla\times(\mathbf{j}\times\mathbf{B}) - R\frac{\partial\theta}{\partial s} \tag{2.20}$$

is often called the thermal wind equation. Following Braginsky (1964,1976) we let

$$u_\phi = U_T + U_M + U_G \tag{2.21}$$

so that u_ϕ is the sum of a thermal wind, a magnetic wind and a geostrophic wind respectively. The thermal and magnetic winds are the solutions of

$$\frac{\partial U_T}{\partial z} = -R\frac{\partial\theta}{\partial s}, \qquad \frac{\partial U_M}{\partial z} = \Lambda\hat{\boldsymbol{\phi}}\cdot\nabla\times(\mathbf{j}\times\mathbf{B}), \tag{2.22}$$

while the geostrophic wind U_G is an arbitrary function of s only. In principle, the thermal wind should be found from a convection calculation, like the α-effect, but in practice some rather arbitrary form is usually assumed.

The meridional circulation

$$\mathbf{u}_\mathrm{p} = \left(-\frac{1}{s}\frac{\partial\psi}{\partial z}, 0, \frac{1}{s}\frac{\partial\psi}{\partial s}\right)$$

is determined by the ϕ-component of the momentum equation. How-

ever, as we noted in Section 3, we must be careful about ignoring viscosity in the ϕ-component, so we retain this term (although it is neglible in (2.20)), to give

$$-\frac{\partial \psi}{\partial z} = \Lambda \mathbf{j} \times \mathbf{B} \cdot \hat{\boldsymbol{\phi}} + E \nabla^2 u_\phi \tag{2.23}$$

Integration of (2.23) over the Taylor cylinders gives us (2.6), when the Ekman boundary layers are taken into account: ψ must vanish at the core-mantle boundary.

We can write the magnetic field in terms of an azimuthal part and a poloidal part, so that

$$\mathbf{B} = \mathbf{B}_\mathrm{p} + B\hat{\boldsymbol{\phi}} = \boldsymbol{\nabla} \times A\hat{\boldsymbol{\phi}} + B\hat{\boldsymbol{\phi}} \tag{2.24}$$

and the induction equations for A and B are

$$\frac{\partial A}{\partial t} + \frac{1}{s}(\mathbf{u}_\mathrm{p} \cdot \boldsymbol{\nabla})(sA) = \alpha B + \left(\nabla^2 - \frac{1}{s^2}\right) A, \tag{2.25}$$

$$\frac{\partial B}{\partial t} + s(\mathbf{u}_\mathrm{p} \cdot \boldsymbol{\nabla})\left(\frac{B}{s}\right) = s(\mathbf{B}_\mathrm{p} \cdot \boldsymbol{\nabla})\left(\frac{u_\phi}{s}\right) - s\boldsymbol{\nabla} \cdot \left(\frac{\alpha}{s^2} \boldsymbol{\nabla}(As)\right) + \left(\nabla^2 - \frac{1}{s^2}\right) B. \tag{2.26}$$

In the $\alpha\omega$ limit these equations can be simplified somewhat. The essence of the $\alpha\omega$ idea is that toroidal velocities are much greater than the meridional circulation. Since the meridional circulation $\mathbf{u}_\mathrm{p}$ is of order η/d, if we take the magnitude of U_T to be $\mathcal{U}_0$, then the magnetic Reynolds number is large,

$$R_m = \frac{\mathcal{U}_0 d}{\eta} \gg 1,$$

in the $\alpha\omega$ limit. This scaling is different from that assumed earlier, in which azimuthal as well as meridional velocities were scaled on η/d. If an $\alpha\omega$ type scaling is to emerge from the convection model, it is necessary that one of the convective parameters has an extreme value, such as large (modified) Rayleigh number R or small q $(= \kappa/\eta)$. The scalars A and B are no longer the same order of magnitude, but $A \sim R_m^{-\frac{1}{2}}$ and $B \sim R_m^{\frac{1}{2}}$, so that the toroidal field is larger than the poloidal field by a factor R_m. This means we can omit the α-effect term in the B equation; toroidal field generation by the α-effect is swamped by toroidal field generation from the large shear. Braginsky (1978) and Braginsky and Roberts (1987) have looked at $\alpha\omega$-dynamo models numerically in spherical geometry, and Proctor (1977) and Ierley (1985) have looked at α^2 models also in spherical geometry. We shall return to a discussion of their results,

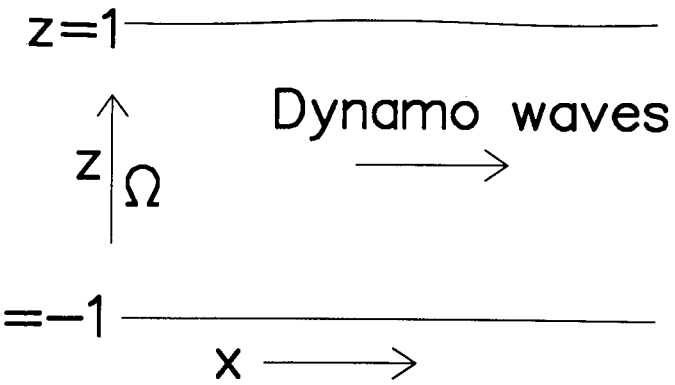


Figure 2.2. Infinite plane layer $\alpha\omega$-model.

but we start by looking at models using simple plane layer and duct geometries which are usually much simpler to solve than spherical models, and so can be used effectively to elucidate the mathematical processes without getting bogged down in numerical difficulties.

2.4.2 Plane Layer and Duct Models

Dynamo models in an infinite slab between parallel planes $z = 0$ and $z = d$ are sometimes considered as models of the galactic dynamo (e.g. Parker, 1979), but they can also be used to elucidate nonlinear behaviour in rapidly rotating dynamo problems. Soward and Jones (1983) considered α^2 dynamos and the $\alpha\omega$ case was studied by Abdel-Aziz and Jones (1988).

In this geometry, illustrated in the $\alpha\omega$ case in Figure 2.2, significant simplifications are possible. We start with the $\alpha\omega$ case, as this is slightly simpler and the analysis has been taken a stage further. A number of different choices for the distribution of α and U_T have been considered, but we illustrate with the simple choice $\alpha = \alpha_0 z$ and $U_T = \frac{1}{2}\mathcal{U}_T z^2$ where z is the dimensionless depth coordinate: the layer boundaries are at $z = \pm 1$ (see Figure 2.2), and the region outside the layer is perfectly insulating. We can divide solutions into those with B symmmetric and A antisymmetric about $z = 0$ (quadrupole parity) or B antisymmetric and A symmetric (dipole parity): we are envisaging that $z = 0$ corresponds to the equatorial plane in spherical geometry. We only describe the behaviour of the dipole modes here. For this model, the induction equations can be written

$$\frac{\partial A}{\partial t} + \frac{\partial(\psi, A)}{\partial(x, z)} = D\alpha B + \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right) A, \tag{2.27}$$

$$\frac{\partial B}{\partial t} + \frac{\partial(\psi, B)}{\partial(x,z)} = z\frac{\partial A}{\partial x} - \frac{\partial A}{\partial z}\frac{\partial U_G}{\partial x} + \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right)B. \qquad (2.28)$$

In this plane geometry there is no magnetic wind in the $\alpha\omega$ case. The dynamo number $D = R_\alpha R_m$, where $R_\alpha = \alpha d/\eta$ is the dimensionless strength of the α-effect. The dynamo number measures the strength of the dynamo action; it is a product of the α-effect and the thermal wind. If ψ and U_G are set to zero, kinematic solutions of (2.27) and (2.28) at $D = D_c$ correspond to travelling wave solutions in the x-direction of the layer. Reversing the sign of the dynamo number reverses the direction of propagation of the dynamo waves, but does not lead to an essentially different solution: note that this property is not carried over to spherical geometry, where changing the sign of the dynamo number leads to totally different solutions.

The y-component of the equation of motion then reduces to

$$\epsilon^{\frac{1}{2}} U_G = -\frac{\partial}{\partial x}\int_{-1}^{1} B\frac{\partial A}{\partial z}\, dz \qquad (2.29)$$

where ϵ is the coupling parameter, which is a renormalised Ekman number in the viscous coupling case. As in the convection problem, if ϵ is small we have a coupling regime when A and B are $O(\epsilon^{\frac{1}{4}})$, and ϵ can then be scaled out. The circulation is determined by (2.23), but in the coupling regime it is $O(\epsilon^{\frac{1}{2}})$ and therefore negligible; the only nonlinearity is therefore that in (2.29). In the infinite plane layer we can look for travelling waves of the form $f(z, x-ct)$ which are periodic in the x-direction with wavenumber k. The wavespeed c comes out as a nonlinear eigenvalue, so the system (2.27) to (2.29) can be solved as a two-dimensional boundary value problem: no time-stepping is necessary, and straightforward spectral methods give accurate solutions.

As D is increased above D_c, the magnitude of the solution increases rapidly as D approaches the second critical dynamo number D_T, at a rate proportional to $(D_T - D)^{-\frac{1}{2}}$ (see Figure 2.3a). This behaviour is analogous to that which occurs in the convection problem: A and B are adjusting their shape so that the Taylor constraint holds: U_G remains constant, while the magnitude of A and B rapidly increase. Once $D > D_T$, we are in the Malkus-Proctor dynamo regime, and the equations we must solve are different. A and B are now $O(1)$ and equation (2.29) becomes

$$\frac{\partial}{\partial x}\int_{-1}^{1} B\frac{\partial A}{\partial z}\, dz = 0, \qquad (2.30)$$

and the meridional circulation is now $O(1)$, and is given by (2.23). The nonlinearity now comes from the meridional circulation, but the

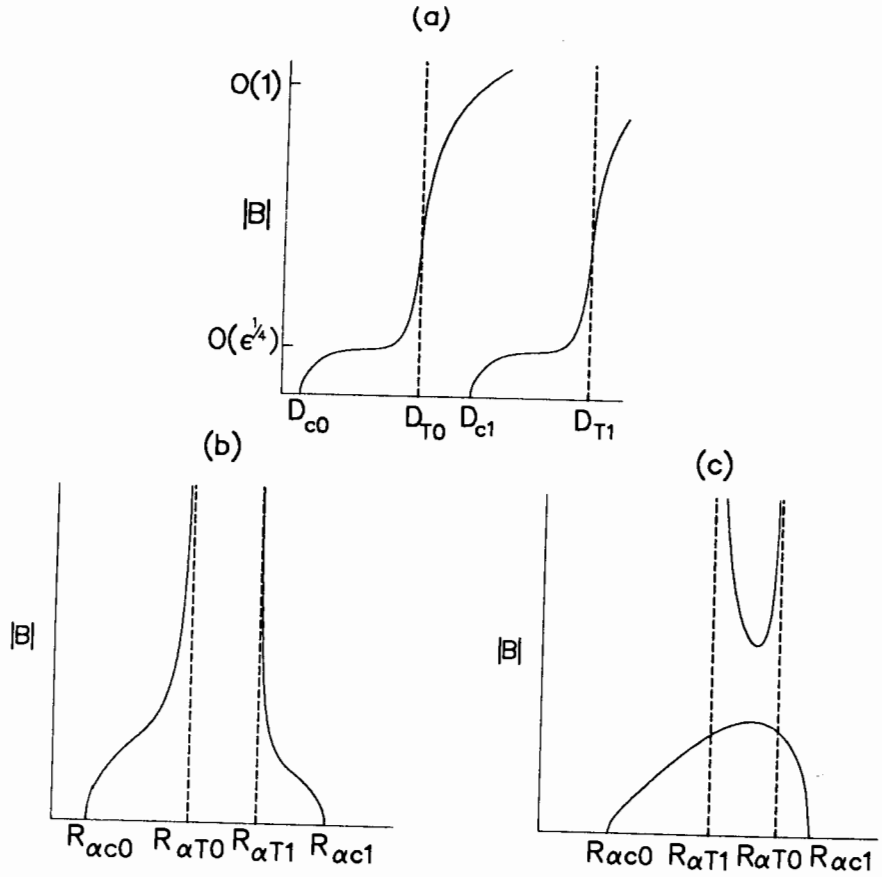


Figure 2.3. Schematic diagrams of the field strength $|B|$ with dynamo number. D_{c0} and D_{c1} are the lowest two dynamo numbers arising from the linear eigenvalue problem. D_{T0} and D_{T1} are the corresponding Taylor points. (a) is for the $\alpha\omega$ case: in (b) and (c) only the coupling regime for the α^2 case is considered, and R_α replaces D.

same numerical techniques can be used. The resulting solutions are independent of core-mantle coupling and are of the type envisaged by J.B. Taylor and Malkus and Proctor. Although the magnitude of the solutions undergoes a sudden jump as we pass from the coupling regime into the Taylor regime (Figure 2.3a), the wavespeed and the geostrophic wind vary continuously across D_T; the dynamo starts with travelling dynamo waves at $D = D_T$ and is still a travelling wave

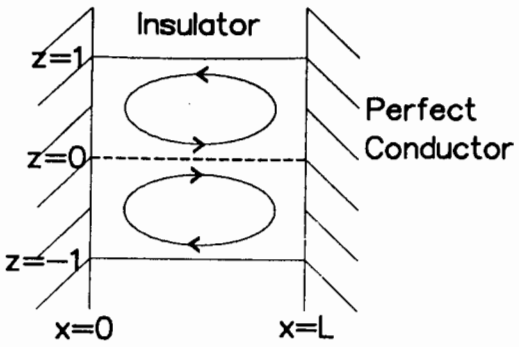


Figure 2.4. The infinite layer of Figure 2.2 becomes a duct. A meridional circulation can be added.

dynamo at large D. It will be noted that on Figure 2.3a a second critical dynamo number D_{c1} and corresponding Taylor number D_{T1} are shown. In general, there is infinite sequence of critical dynamo numbers and corresponding Taylor states; we have just marked the first two.

In investigating the α^2 dynamo problem in the same geometry, Soward and Jones (1983) found that other ways of connecting the solutions could occur: in Figure 2.3b we show schematic sketches of the connections found for various distributions of the α-effect. There is no thermal wind in α^2-dynamos, so R_α plays the role of the dynamo number. Note that in Figure 2.3b the second linear critical R_α comes after the corresponding solution, so this bifurcation is subcritical rather than supercritical. In Figure 2.3c another phenomenon is shown: the Taylor solutions are disconnected from the solutions bifurcating out of the linear critical solutions. It is possible to vary a parameter so that the α-distribution moves smoothly from case 3b to case 3c; there is then a critical value at which the reconnection of the solution takes place.

The plane layer models are much simpler to handle than models in a bounded geometry, but the $\alpha\omega$ models have the unsatisfactory feature that the dynamo waves progress for ever, whereas in a realistic problem they are blocked at the boundaries. In order to investigate the effect of blocking, Wallace (1990) has put vertical conducting sidewalls on the infinite plane layer, thus turning it into a duct (see Figure 2.4). This problem has to be solved by time-stepping and it forms a useful preliminary to doing the problem in a full sphere.

The equations solved are (2.27) and (2.28) with $\psi = 0$, together

with (2.29). The integration of the equations is not as straightforward as we might imagine: we can see how the difficulty arises when we ask the question why the geostrophic wind shear cannot sustain the dynamo. As we would expect, the answer comes from the energy integral: the dynamo is convection-driven and only the thermal wind can increase toroidal magnetic energy. We rescale A and B with the factor $\epsilon^{\frac{1}{2}}$, multiply (2.28) by B and integrate over the whole area of the duct S to obtain

$$\frac{\partial}{\partial t}\int_S \frac{1}{2}B^2\,ds = \int_S zB\frac{\partial A}{\partial z}\,ds - \int_0^L \Big(\frac{\partial}{\partial x}\int_{-1}^{1} B\frac{\partial A}{\partial z}\,dz\Big)^2\,dx - \int_S |\nabla B|^2\,ds. \tag{2.31}$$

Various integrated terms vanish because of the boundary conditions. The second term on the right is the geostrophic wind term, and it is clear that this can only be dissipative, i.e. negative definite (Childress, 1969). The term is therefore a nonlinear diffusion term, which makes it awkward numerically. The numerical scheme must preserve this negative definite property for all possible distributions of A and B, or else severe numerical instability may occur. In practice, the Galerkin technique is the only one so far discovered to be entirely stable. This technique involves taking an integration over space, so that the form of (2.31) is adhered to: the vital step is to ensure that although the finite truncations representing A and B are only approximate, the matrices produced are still exactly symmetric and negative definite.

Despite the blocking of the dynamo waves, these $\alpha\omega$ dynamos are still oscillatory, but now they represent standing, rather than travelling waves. $\alpha\omega$ dynamos can give steady modes in some circumstances, but even when the first mode to onset as D is increased is a steady mode, it is generally found that oscillatory modes rapidly take over as D is increased beyond D_c. The steady $\alpha\omega$ dynamo is a very delicate creature (Parker 1979). Having said this, there is one way an $\alpha\omega$ dynamo can be made steady in a reasonably robust way, and that is by adding meridional circulation (Roberts, 1972). The meridional circulation does have to be of an appropriate strength and direction, however. The meridional circulation found in the Abdel-Aziz and Jones plane layer dynamo did not steady the dynamo: indeed, the wave frequency increased as the circulation strengthened. Nevertheless, although the geodynamo reverses occasionally, there have been times when it has had the same sign for many years, so that a steady model is probably a better "first approximation" than an oscillatory

model. There is no conclusive evidence that the geodynamo is of the $\alpha\omega$ type, but the balance of the available evidence suggests that it is: the observed westward drift suggests toroidal velocites that are substantially higher than the magnetic diffusion velocity scale η/d, so that the magnetic Reynolds number is probably at least 100. Also, a toroidal field of a few hundred gauss, which is consistent with $R_m \sim 10^2$, would make the Elsasser number of order 1, which is believed to be optimal for convection (see Section 3). If we accept this as evidence that the geodynamo contrives to be a steady $\alpha\omega$ dynamo, then it is of interest to examine the effect of adding an imposed meridional circulation to the duct model and seeing what behaviour can result.

A simple meridional circulation of the form (see Figure 2.4)

$$\psi = mx(1-x)z(1-z^2)$$

which preserves the dipole-quadrupole symmetry of the solution is chosen, and equations (2.27) to (2.29) solved. This in itself does not achieve the aim of producing steady solutions, although some rather complex time-dependent behaviour is found. In addition, the thermal wind U_T is generalised to the form

$$U_T = \frac{1}{2}\mathcal{U}_z z^2 - \mathcal{U}_x x$$

so that the term $z\partial A/\partial x$ in (2.28) is generalised to $\mathcal{U}_z z\partial A/\partial x + \mathcal{U}_x \partial A/\partial z$. This allows a wide range of possible nonlinear behaviour, which we cannot elucidate fully here.

In Figures 2.5a,b,c and d we show some of the types of solution found. In Figure 2.5a, $m = \mathcal{U}_x = 0$, the case most closely resembling the infinite layer solutions, we show the amplitude of the lowest spectral function representing A, which is a convenient measure of the amplitude of the solution. After an initial transient the solution settles down to a steady periodic oscillation about zero mean. Figure 2.5b has m non-zero, but $\mathcal{U}_x = 0$. The dynamo number is also somewhat larger. The solution appears to be chaotic, never settling down to periodic or quasiperiodic behaviour. This type of behaviour is known in other types of nonlinear dynamo (see e.g. Weiss, Cattaneo and Jones, 1984) where rapid rotation is not involved. Figure 2.4c is a steady dynamo produced with non-zero m and both x and z shear; in fact $\mathcal{U}_x = \mathcal{U}_z = 1$ for this run. The dynamo number is considerably lower in this case than for cases 2.5a and 2.5b: reduction of the critical dynamo number with meridional circulation was also noted by Roberts (1972). Case 2.5d also has $\mathcal{U}_x = \mathcal{U}_z = 1$ and non-zero m,

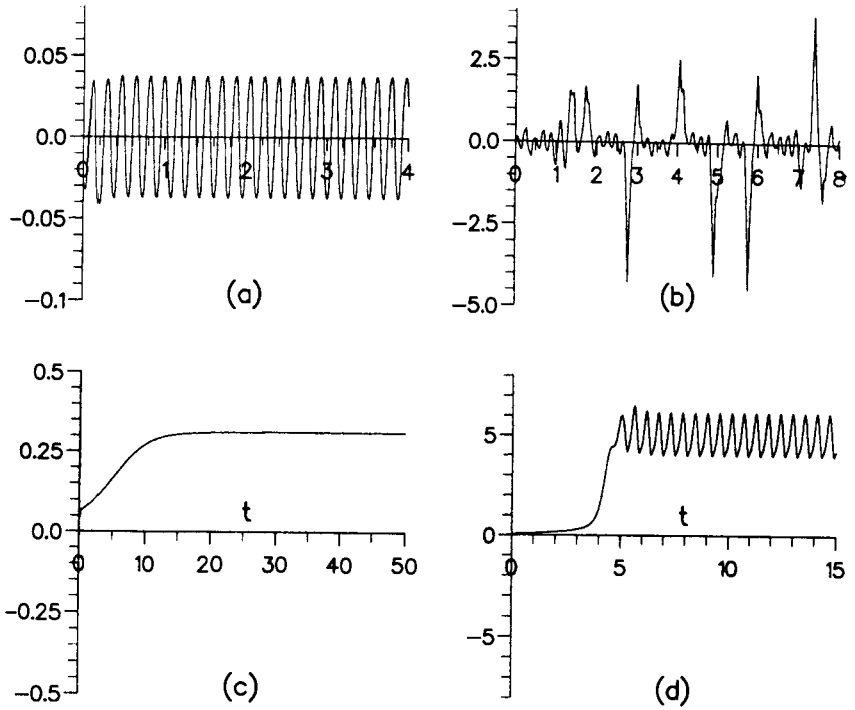


Figure 2.5. Field strength plotted against time. (a) D=2500, $\mathcal{U}_x$=0, $\mathcal{U}_z$=1, m=0. (b) As (a), but with D=3000 and m=50. (c) D=350, $\mathcal{U}_z = \mathcal{U}_x = 1$, $m = -40$. (d) As (c) but with D=700 and m=-25.

but here m is smaller than in case (c), and the tendency to oscillate has not been eliminated. The resulting oscillation about non-zero mean is sometimes called vacillation. It is quite an attractive type of solution for modelling the geodynamo, which is known from palaeomagnetic data to vary significantly even when it does not undergo a complete reversal.

All these solutions assume the dipole symmetry: we know from other nonlinear dynamo studies (Jennings and Weiss 1990) that if

the nonlinear interactions between quadrupole and dipole modes are considered, the picture can become even more complicated.

2.4.3 Spherical Models

A number of authors have investigated rapidly rotating dynamo models in a spherical geometry. Here we discuss briefly Braginsky's model-Z (1976, 1978, see also Braginksy and Roberts 1987) and the Malkus-Proctor (1975) model developed by Proctor (1977) and Ierley (1985). These models all adopted the time-stepping approach and took the Ekman number as finite but numerically small.

The model-Z dynamo is of $\alpha\omega$ type. The imposed thermal wind used is not very special, but the α-distribution is concentrated in a narrow strip in the equatorial belt near the core-mantle boundary. This distribution, together with a meridional circulation satisfying equation (2.23) gives rise to a steady dynamo. This solution does not satisfy Taylor's condition in the way described above, but behaves in a different manner: the Taylor integral for axisymmetric fields has the form

$$\int_S BB_s \, dz$$

B_s being the component of the poloidal field in the s-direction. In the model-Z solutions, B_s is very small compared to B_z, the component of meridional field in the rotation direction. This reduces the value of the magnetic torque, and permits a balance with the weak coupling terms. The upshot of this is that everywhere in the interior of the core the poloidal field is almost parallel to the rotation axis, hence the name model-Z. The poloidal field changes direction in a thin layer near the core-mantle boundary. The geostrophic wind is related to the magnetic torque though equation (2.6); although the Taylor integral is reduced in magnitude because B_s is small, U_G is nevertheless still quite large. The scalings which model-Z apparently tends to as $E \to 0$ are discussed in Soward (1989). The relationship of model-Z to Malkus-Proctor dynamos is discussed in Roberts (1989). Model-Z is a successful dynamo model, but it does have some drawbacks. The special concentrated distribution of α seems to be an essential feature of the model, and attempts to replace this with more uniform distributions have not produced convergent solutions. The question of how the solutions are connected with kinematic solutions is also not yet resolved. Malkus and Proctor (1975) suggested approaching the problem by increasing D from kinematic values until a Taylor state is achieved: this motivated the plane layer studies in

which their goals have been realised. Proctor (1977) studied the α^2 dynamo problem numerically using a finite difference technique and concluded that there was evidence for the existence of a transition to a Taylor state at a particular value of the dynamo number. These calculations were taken futher by Ierley (1985) who came to a similar conclusion. Ierley did find, however, that even at large D the solution did still depend on the Ekman number, so that a genuine Taylor regime appears not to have been reached. This may be because it was not possible to reduce the Ekman number to sufficiently low values: numerical schemes using finite Ekman number often encounter numerical instability at very low E. Ierley also examined a number of different α-distributions, and found that some of them were consistent with the idea that the Taylor states were disconnected from the branch emanating from the linear critical dynamo number, as in Figure 2.3c rather than in Figure 2.3b.

There are clearly still many unresolved problems even in the axisymmetric dynamo problem concerning the existence of Taylor states and the status of model-Z. Work is in progress to attempt to resolve some of these problems (Barenghi, 1990).

A different approach to the problem was suggested by Fearn and Proctor (1984, 1987). In an ambitious attempt to eliminate the arbitrariness of choosing α-distributions and thermal wind, they proposed an iterative scheme in which the linear nonaxisymmetric magnetoconvection problem was solved to provide estimates for the mean e.m.f. $\overline{\mathbf{u}' \times \mathbf{b}'}$ to be used in the axisymmetric dynamo problem. The hope was that the fields generated by the solutions of the dynamo problem could be returned to the magnetoconvection problem and a convergent sequence of α-distributions obtained. Although much was learned about magnetoconvection and axisymmetric dynamo models en route, the iteration scheme never converged. The reasons for this are not clear, but the difficulty seemed to be in finding solutions which satisfy Taylor's constraint.

Acknowledgements

The author is grateful for much helpful advice from Professor A.M. Soward and Dr. C.F. Barenghi and also thanks Mr. S.G. Wallace for assistance with numerical calculations. He has also benefitted from discussions with Professor P.H. Roberts, Dr. D.R. Fearn and Dr. M.R.E. Proctor. Some material in this chapter derives from work done under SERC research grant GR/E93251.

References

Abdel-Aziz, M.M. and Jones, C.A. 1988. $\alpha\omega$-dynamos and Taylor's constraint. *Geophys. Astrophys. Fluid Dynam.*, **44**, 117-139.

Acheson, D.J., 1983. Local analysis of thermal and magnetic instabilities in a rapidly rotating fluid. *Geophys. Astrophys. Fluid Dynam.*, **27**, 123-136.

Barenghi, C.F., 1990. In preparation.

Braginsky, S.I., 1964. Theory of the hydromagnetic dynamo. *JETP*, **47**, 2178-2193.

Braginsky, S.I., 1975. An almost axially symmetric model of the hydromagnetic dynamo of the Earth I. *Geomagn. Aeron.*, **15**, 149-156.

Braginsky, S.I., 1976. On the nearly axially symmetric model of the hydromagnetic dynamo of the Earth. *Phys. Earth Planet. Inter.*, **11**, 191-199.

Braginsky, S.I., 1978. An almost axially symmetric model of the hydromagnetic dynamo of the Earth II. *Geomagn. Aeron.*, **18**, 240-351.

Braginsky, S.I. and Roberts, P.H., 1987. A model-Z geodynamo. *Geophys. Astrophys. Fluid Dynam.*. **38**, 327-349.

Busse, F.H., 1975. A model of the geodynamo. *Geophys. J. R. Astr. Soc.*, **42**, 437-459.

Busse, F.H., 1976. Generation of planetary magnetism by convection. *Phys. Earth Planet. Inter.*, **12**, 350-358.

Childress, S., 1969. A class of solutions of the magnetohydrodynamic dynamo problem. In: *The application of Modern Physics to the Earth and Planetary Interiors* ed. S.K.Runcorn, Wiley, 629-648.

Cowling, T.G., 1957. The dynamo maintenance of steady magnetic fields. *Q.J. Mech. Appl. Math.*, **10**, 129-136.

Cupal, I., and Hejda, P., 1989. On the Computation of a Model-Z with Electromagnetic Core-Mantle Coupling. *Geophys. Astrophys. Fluid Dynam.*, **49**, 161-172.

Eltayeb, I.A., 1972. Hydromagnetic convection in a rapidly rotating fluid layer. *Proc. R. Soc. Lond.*, **A326**, 229-254.

Eltayeb, I.A., 1975. Overstable hydromagnetic convection in a rapidly rotating fluid layer. *J. Fluid Mech.*, **71**, 161-179.

Eltayeb, I.A. and Kumar, S., 1977. Hydromagnetic convective instabilities of a rotating self-gravitating fluid sphere containing a uniform distribution of heat sources. *Proc. R. Soc. Lond.*, **A353**, 145-162.

Fearn, D.R., 1979a. Thermally driven hydromagnetic convection in a rapidly rotating sphere. *Proc. R. Soc. Lond.*, **A369**, 227-242.

Fearn, D.R., 1979b. Thermal and magnetically driven instabilities in a rapidly rotating fluid sphere. *Geophys. Astrophys. Fluid Dynam.*, **14**, 103-126.

Fearn, D.R., 1983. Hydromagnetic waves in a differentially rotating annulus I. A test of local stability analysis. *Geophys. Astrophys. Fluid Dynam.*, **27**, 137-162.

Fearn, D.R., 1989. Compositional convection and the Earth's core. In: *Geomagnetism and Palaeomagnetism* eds. F.J. Lowes, D.W. Collinson, J.H. Parry, S.K. Runcorn, D.C. Tozer and A.M. Soward, NATO ASI series C, Vol 261, 335-346.

Fearn, D.R. and Proctor, M.R.E., 1983. Hydromagnetic waves in a differentially rotating sphere. *J. Fluid Mech.*, **128**, 1-20.

Fearn, D.R. and Proctor, M.R.E., 1984. Self-consistent dynamo models driven by hydromagnetic instabilities. *Phys. Earth Planet. Inter.*, **36**, 78-84.

Fearn, D.R. and Proctor, M.R.E., 1987. On the computation of steady, self-consistent, spherical dynamos. *Geophys. Astrophys. Fluid Dynam.*, **38**, 293-325.

Fearn, D.R., Roberts, P.H. and Soward, A.M., 1988. Convection, stability and the dynamo. In: *Energy, Stability and Convection* eds G.P. Galdi and B. Straughan, Pitman Research Notes in Mathematics series, Longmans, Vol 168, 60-324.

Fearn, D.R., and Weiglhofer, W.S., 1990. Magnetic instabilities in rapidly rotating spherical geometries I. From cylinders to spheres. Submitted to *Geophys. Astrophys. Fluid Dynam.*

Gubbins, D. and Roberts, P.H., 1988. Magnetohydrodynamics of the Earth's core. In: *Geomagnetism* Vol. 2 (J.A. Jacobs, ed.), Academic Press, London, 1-183.

Hide, R., 1966. Free hydromagnetic oscillations of the Earth's core and the theory of the geomagnetic secular variation. *Phil. Trans. R. Soc. Lond.*, **A259**, 615-650.

Ierley, G.R., 1985. Macrodynamics of α^2-dynamos. *Geophys. Astrophys. Fluid Dynam.*, **34**, 143-173.

Jennings, R.L. and Weiss, N.O., 1990. Symmetry breaking in the Solar Dynamo. In: *Solar Photosphere: Structure, Convection and Magnetic Fields* J.O.Stenflo, ed. Reidel, Dordrecht, 355-358.

Jones, C.A., and Roberts, P.H., 1990. Magnetoconvection in rapidly rotating Boussinesq and compressible Fluids. Submitted to *Geophys. Astrophys. Fluid Dynam.*

Jones, C.A., Roberts, P.H., and Galloway, D.J. 1990. Compressible convection in the presence of rotation and a magnetic field. *Geophys. Astrophys. Fluid Dynam.*, **53**, 145-182.

Krause, F. and Rädler, K.-H., 1980. *Mean-Field Magnetohydrodynamics and Dynamo Theory.* Pergamon, Oxford.

Malkus, M.V.R. and Proctor, M.R.E., 1975. The macrodynamics of α-effect dynamos in rotating fluids. *J. Fluid Mech.*, **67**, 417-443.

Moffatt, H.K., 1978. *Magnetic Field Generation in Electrically Conducting Fluids.* Cambridge University Press.

Parker, E.N., 1979. *Cosmical Magnetic Fields, their origin and their activity.* Clarendon Press, Oxford.

Proctor, M.R.E., 1977. Numerical solutions of the non-linear α-effect dynamo equations. *J. Fluid Mech.*, **80**, 769-784.

Roberts, P.H., 1968. On the thermal instability of a rotating fluid sphere containing heat sources. *Phil. Trans. R. Soc. Lond.*, **A263**, 93-117.

Roberts, P.H., 1988a. Origin of the main field: Dynamics. In: *Geomagnetism* Vol. 2 (J.A. Jacobs, Ed.). Academic Press, London, 251-306.

Roberts, P.H., 1988b. Future of geodynamo theory. *Geophys. Astrophys. Fluid Dynam.*, **44**, 3-32.

Roberts, P.H., 1989. From Taylor state to Model-Z? *Geophys. Astrophys. Fluid Dynam.*, **49**, 143-160.

Roberts, P.H. and Gubbins, D., 1988. Origin of the main field: Kinematics. In: *Geomagnetism* Vol. 2 (J.A. Jacobs, ed.) Academic Press, London, 185-249.

Roberts, P.H. and Stewartson, K., 1974. On finite amplitude convection in a rotating magnetic system. *Phil. Trans. R. Soc. Lond.*, **A277**, 287-315.

Roberts, P.H. and Stewartson, K., 1975. Double roll convection in a rotating magnetic system. *J. Fluid Mech.*, **68**, 447-466.

Skinner, P.H. and Soward, A.M., 1988. Convection in a rotating magnetic system and Taylor's constraint. *Geophys. Astrophys. Fluid Dynam.*, **44**, 91-116.

Skinner, P.H. and Soward, A.M., 1990. Convection in a rotating magnetic system and Taylor's constraint II. Submitted to *Geophys. Astrophys. Fluid Dynam.*

Soward, A.M., 1972. A kinematic theory of large magnetic Reynolds number dynamos. *Phil. Trans. R. Soc. Lond.*, **A272**, 431-462.

Soward, A.M., 1986. Non-linear marginal convection in a rotating magnetic system. *Geophys. Astrophys. Fluid Dynam.*, **35**, 329-371.

Soward, A.M., 1989. Geodynamo Theory. In: *Geomagnetism and Palaeomagnetism* eds. F.J. Lowes, D.W. Collinson, J.H. Parry, S.K. Runcorn, D.C. Tozer and A.M. Soward, NATO ASI series C, Vol 261, 297-319.

Soward, A.M. and Jones, C.A., 1983. α^2-dynamos and Taylor's constraint. *Geophys. Astrophys. Fluid Dynam.*, **27**, 87-122.

Taylor, J.B., 1963. The magneto-hydrodynamics of a rotating fluid and the Earth's dynamo problem. *Proc. R. Soc. Lond.*, **A274**, 274-283.

Tough, J.G. and Roberts, P.H., 1968. Nearly axisymmetric hydromagnetic dynamos. *Phys. Earth Planet. Inter.* **1**, 288-296.

Wallace, S.G. 1990. PhD. thesis, University of Newcastle-upon-Tyne (in preparation).

Weiss, N. O., Cattaneo, F., and Jones, C.A., 1984. Periodic and Aperiodic Dynamo waves. *Geophys. Astrophys. Fluid Dynam.*, **30**, 305-341.

Zhang, K.-K. and Busse, F.H., 1988. Finite amplitude convection and magnetic field generation in a rotating spherical shell. *Geophys.Astrophys. Fluid Dynam.*, **44**, 33-54.

Zhang, K.-K. and Busse, F.H., 1989. Convection Driven Magnetohydrodynamic Dynamos in Rotating Spherical Shells. *Geophys. Astrophys. Fluid Dynam.*, **49**, 97-116.

Zhang, K.-K. and Busse, F.H., 1990. Generation of magnetic fields by convection in a rotating spherical fluid shell of infinite Prandtl number. *Phys. Earth Planet. Inter.*, **59**, 208-222.

3

Problems of Planetary Dynamo Theory

F H BUSSE

University of Bayreuth, Germany

3.1 Introduction

One of the most fascinating aspects of the atmospheric dynamics of the planets and of the Sun is the similarity with laboratory experiments in which forces and rotational constraints analogous to those of the celestial bodies have been simulated. In his experiments on convection in thin horizontal layers of liquid heated from below, for instance Henri Bènard was motivated by the analogy with phenomena of solar granulation and of convection in the Earth's atmosphere [Bénard and Avsec, 1938]. The baroclinic waves observed in the dishpan experiments of Fultz (1953) and in the annulus experiments of Hide (1958) have long been recognized as useful simulations of weather systems in mid-latitudes. There is also growing evidence that the band structure of the major planets can be understood in terms of convection columns driven by radial buoyancy in a rotating cylindrical annulus [e.g. Busse, 1983].

The general similarity between the highly turbulent processes on the scales of planets and stars and the laminar or barely turbulent flows in laboratory experiments has traditionally been interpreted in terms of eddy diffusivities (Taylor 1915, Prandtl 1925). While it is generally recognized that this concept is too crude to describe detailed features of the observations, it has been surprisingly successful in providing global estimates of heat and momentum transport in terms of an appropriate mixing length. Laboratory experiments have thus been useful as reliable reference points for theories that attempt to bridge the huge differences in the orders of Reynolds and Rayleigh numbers between the human and the celestial scales.

There are no laboratory experiments available which simulate the generation of planetary or stellar magnetic fields through motions

5

Magnetic Buoyancy

D.W. HUGHES

University of Cambridge, Silver Street, Cambridge CB3 9EW, U.K.

5.1 Introduction

The idea of magnetic buoyancy was conceived by E.N. Parker in 1955, one of his many seminal contributions to solar physics, and independently by E. Jensen in the same year. The term is now used to refer to two different, though related, phenomena. The first concerns the rise of isolated magnetic flux tubes, the original motivation of Parker and Jensen, and is best illustrated by the following simple example. Consider an isolated horizontal flux tube in pressure equilibrium with its non-magnetic surroundings; thus

$$p_i + \frac{B^2}{2\mu_0} = p_e, \tag{5.1}$$

where p_i and p_e are, respectively, the gas pressures internal and external to the tube. If, for simplicity, we assume that the tube and surroundings are in thermal equilibrium then it follows from the equation of state of a perfect gas that, since $p_i < p_e$ then $\rho_i < \rho_e$, i.e. the tube is less dense than its surroundings and will rise under the influence of gravity.

The second rôle played by magnetic buoyancy is as an instability mechanism of a stratified, large-scale magnetic field. In a general field distribution, just as for an isolated flux tube, part of the pressure burden is shouldered by the field. In particular, if the field decreases upwards then the gas is, to a certain degree, top-heavy, and extra gravitational potential energy is available that can be channelled into driving an instability. Such instabilities were first studied by Newcomb (1961), using the energy principle of Bernstein *et al.* (1958), and by Parker (1966), using a normal-mode analysis, although Kruskal & Schwarzschild (1954) had previously considered a related instability of a discontinuous field.

Magnetic buoyancy, in at least one of its guises, is of importance in a number of astrophysical contexts. We shall here be primarily concerned with the magnetic field of the solar interior, but magnetic buoyancy may also play a key rôle in the formation of interstellar gas clouds (e.g. Parker 1969), and in the internal dynamics of accretion discs (e.g. Stella & Rosner 1984). We shall defer a detailed discussion of the implications of magnetic buoyancy for the solar field until §5.4, after having considered the physical ideas involved. Here, in the introduction, we shall just mention briefly why magnetic buoyancy is expected to be an important influence.

Magnetic flux is often observed on the surface as bipolar regions, such as sunspots, aligned roughly parallel to the equator. This strongly suggests the eruption through the photosphere is of predominantly toroidal flux tubes — the behaviour of such buoyant tubes is discussed in §5.2. Whereas the nature of the surface field can be observed directly, the form of the field in the interior can only be inferred theoretically. Although there are several differing theories concerning the sub-surface field, a common feature of many is that, in order to supply the flux observed at the surface there must exist, either in the convection zone or possibly in the underlying region of convective overshoot, a substantial predominantly toroidal field. Such a field is susceptible to instabilities driven by magnetic buoyancy — it is these that are the topic of §5.3.

5.2 The Magnetic Buoyancy of Flux Tubes

High resolution observations of the solar photosphere reveal magnetic flux to be concentrated into intense regions, over a considerable range of scales — from sunspots with diameters of the order of 20,000 km right down to magnetic knots, smaller (in diameter) by a factor of one hundred, at the current limit of spatial resolution. As noted above, the emergence of sunspots as bipolar pairs aligned roughly parallel to the equator provides convincing evidence of toroidal flux tubes bursting through the surface, such tubes presumably being strongly influenced by the effects of magnetic buoyancy. In order that we may eventually answer fundamental questions concerning the location of the bulk of the solar magnetic field, and the mechanism of its regeneration, it is essential to understand the dynamics of flux tubes rising through the convection zone.

Although observations can only tell us the strength of the magnetic field when it emerges at the surface, it is possible, by making

certain assumptions, to obtain a rough estimate of the field strength in the sub-surface layers. From observations of the flux emerging in active regions, together with assumptions about how green the Sun is in recycling used flux, Galloway & Weiss (1981) estimated the total (unsigned) flux in the convection zone (or just below) at sunspot maximum to be $O(10^{24})$Mx (Parker 1987 reaches a similar conclusion from interpreting more recent observations). Although the field strength arising from this total flux depends of course on how tightly squashed is the field, it does appear that the field strength will be quite high, with magnetic energy comparable to the kinetic energy density. Galloway & Weiss postulate that the field is stored in the region of convective overshoot below the convective zone proper and has a strength of 10^4G; Parker, on the other hand, suggests that the field is stored in the lower half of the convective zone with a strength of about 3×10^3G. In §5.4 we shall return to the issue of where, and in what form, the solar field might be stored.

Describing the dynamics of flux tubes rising through the differentially rotating, turbulent convective zone is obviously extremely difficult and certain simplifications have to be made in order to obtain a tractable problem. The most widely used simplification is to ignore all interactions between large-scale convective motions and the rising flux tubes. Typically it is also assumed that the flux tubes are thin (i.e. with radius much smaller than local scale heights) so that variations across the tubes may be neglected. If, further, a rising tube is assumed to remain horizontal then its evolution, being governed by its buoyancy and any retarding drag force, is reasonably straightforward. The mass and magnetic flux of the tube are conserved and the tube adjusts rapidly, on the fast magneto-acoustic timescale, to being in total pressure equilibrium with its surroundings. The buoyancy force on the tube is then determined once the thermodynamic behaviour is known — for example, whether the tube expands adiabatically, adjusts instantaneously to the surrounding temperature, or acts somewhere in between these extremes.

Parker (1975) estimated the rate of rise of a horizontal flux tube by balancing its buoyancy against its aerodynamic drag, in a uniform atmosphere. He concluded that magnetic buoyancy was rather efficient and suggested that flux tubes with field strengths as low as 100G would be removed from the convection zone in a couple of years. Parker's calculation was extended, still for purely horizontal tubes, by a number of authors. Unno & Ribes (1976) incorporated a large damping term to model eddy viscosity due to turbulence out-

side the tube, the effect being to decrease the upward velocity of the tube by a factor V_A/V_S (V_A = Alfvén speed, V_S = sound speed). Schüssler (1977) considered the rise of a tube always in thermal equilibrium with its surroundings, taking into account the stratification of the external medium and the resulting expansion of the tube as it rises. The concomitant reduction in the buoyancy of the tube is found to be rather large — a factor of 20 compared to Parker's estimate, when the same retarding force is assumed. The most comprehensive study of the behaviour of buoyant, horizontal flux tubes is by Moreno-Insertis (1983), who looked at a variety of cases. The most significant result of his work is that the thermal behaviour of the tube is unimportant for tubes with high field strength ($\gtrsim 10^4 G$), with adiabatic and isothermal tubes rising almost identically, whereas for tubes of smaller field strengths, those expanding adiabatically rise considerably faster (a factor of 100 when $B \sim 100$G) than those that adjust instantaneously to the external atmosphere. This is due to the fact that adiabatic tubes retain their inherent buoyancy in a convectively unstable medium (regardless of the magnetic field), but this is annulled for tubes that attain the same temperature as their surroundings. Moreno-Insertis, as did Schüssler, considered the braking effects of both aerodynamic drag and turbulent viscosity. The latter was found to be the more effective retarding force — even so, flux tubes of equipartition field strength were found to rise through the convection zone in a matter of months.

It is important to note that the problem of rising flux tubes is typically *not* one of stability. From (1) it follows that only for one special value of ΔT, the temperature difference between the external medium and the tube, $\Delta T = B^2 m/2\mu_0 k_B \rho = T_c$, say, will the tube be in mechanical equilibrium (with $\rho_i = \rho_e = \rho$). If $\Delta T < T_c$ then $\rho_i < \rho_e$ and the tube is buoyant; if for some reason the tube is so cool that $\Delta T > T_c$ then the magnetic buoyancy will be overwhelmed and the tube will sink. The stability of the special case when $\Delta T = T_c$ has been studied by Spruit & van Ballegooijen (1982). Tubes displaced bodily, without bending, are unstable only if

$$\delta > (1 - \gamma/2)/(\gamma(1 + \beta\gamma/2)), \qquad (5.2)$$

where β is the ratio of the gas pressure inside the tube to the magnetic pressure, and δ (defined as $d\ln T_e/d\ln p_e - 1 + 1/\gamma$) is positive (negative) if the stratification of the external medium is superadiabatic (subadiabatic). It can be seen from (5.2) that instability can only set in for convectively unstable atmospheres ($\delta > 0$) and that

increasing the field strength (decreasing β) has a stabilising effect. The difference in the behaviour of buoyant flux tubes, such as those in thermal equilibrium with their surroundings, and that of tubes displaced adiabatically from mechanical equilibrium is striking. For the former, larger field strengths lead naturally to enhanced buoyancy; for the latter, a parcel argument reveals that a strong field reduces the expansion of a tube displaced upwards, thus making it more stable than in the complete absence of field. Of course, since in the latter case the initial premise is that the tube has the same density as its surroundings, the indirect effect of the magnetic field is to cool the tube to the requisite temperature for mechanical equilibrium.

All of the above calculations have neglected the effects of solar rotation. However, since the time taken for even the most rapidly rising flux tube to traverse the convection zone is of the order of a month, comparable to a solar day, the effects of rotation could be important. The influence of rotation on the rise of isolated flux tubes has been considered by Choudhuri & Gilman (1987), who studied axisymmetric tubes rising through a uniformly rotating convection zone. (Acheson 1979a also noted the possible significance of rotation, but was principally concerned with the instability of a large-scale field (see §5.3).) Both magnetic tension and centrifugal forces turn out to be unimportant, the motion of the tubes being determined by buoyancy, Coriolis and drag forces. Rather interestingly, it is found that for tubes with $B \sim 10^4 G$ the outward (i.e. away from the rotation axis) component of the buoyancy force is quickly cancelled by the inward Coriolis force, the net effect being that the tubes rise parallel to the rotation axis. Consequently, even tubes that start their ascent in equatorial regions pop out at the surface at high latitudes; only for tubes of greater field strength will the magnetic buoyancy be sufficiently large as to produce an almost radial rise. As pointed out by Choudhuri & Gilman, the fact that their model behaves very differently to the Sun – where large active regions appear only in mid- to low-latitudes – suggests that the non-axisymmetry of the tubes, with possible anchoring at depth, is a significant feature.

Once the artificial constraint of axisymmetry (or the tube always remaining horizontal) is removed, the dynamics inevitably becomes more complicated. Not only will the tube again adjust rapidly (if it is thin) to having the same total pressure as the adjacent gas, but will also strive to attain hydrostatic equilibrium along its length. The ensuing axial flow is typically, though not always, away from

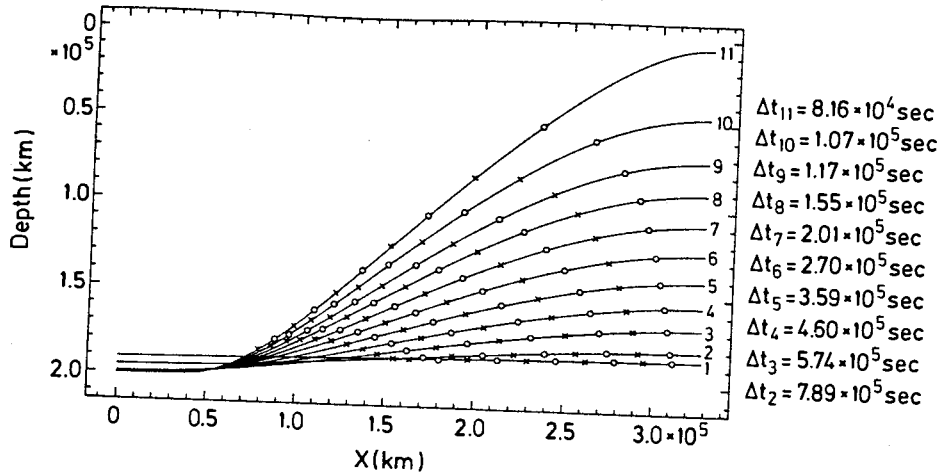


Figure 5.1. Time evolution of a kink-unstable flux tube with initial field strength of $10^{5.5}$G, flux of 10^{21}Mx, and at an initial depth of 1.9×10^5km, whose shape has been given a small-amplitude sinusoidal perturbation in the vertical plane. Time intervals between consecutive curves are shown on the right. Circles and crosses indicate fixed Lagrangian elements of the flux tube (from Moreno-Insertis 1986).

the crests of the bent tube and into the troughs. If this is the case then, as pointed out by Parker (1955) in his original paper on the subject, the crests are drained of mass, their cross-sectional area decreases and hence their field strength and buoyancy are enhanced (the opposite process occurring in the troughs).

A similar conclusion was reached by Spruit & van Ballegooijen (1982) in their study of the stability of a tube in mechanical equilibrium. They found that instability to wavy modes sets in if

$$k_y^2 < \frac{(1+1/\beta)}{2H_p^2}(1/\gamma + \beta\delta), \tag{5.3}$$

where k_y is the wavenumber along the tube and H_p is the pressure scale height of the external medium. In contrast to modes that raise the entire tube, with instability criterion (5.2), instability can now occur in convectively stable atmospheres ($\delta < 0$). In this sense the field may be thought of as being destabilising, although it should be pointed out that for convectively unstable atmospheres ($\delta > 0$), increasing the strength of the field is a stabilising influence, just as for the simpler modes discussed above. For a very weak field ($\beta \gg 1$) the growth rate of the instability is proportional to $\delta^{1/2}$, whereas for a very strong field ($\beta \ll 1$) the growth rate is reduced to $O(\beta^{1/2})$.

The most ambitious study to date of the evolution of isolated flux

tubes is that of Moreno-Insertis (1986). He considered tubes, initially horizontal, lying in mechanical equilibrium ($\rho_i = \rho_e$) towards the bottom of the convection zone — the field strengths being such as to allow instability to certain undulatory perturbations, but to forbid instability due to bodily displacements of the tube as a whole (c.f. inequalities (5.2) and (5.3)). The ascent was assumed to be sufficiently rapid that it could be regarded as occurring adiabatically. The rise of the buoyant crests is accompanied by a persistent downflow along the tube. As explained earlier, this enhances the buoyancy of the crests; it also leads to an increase in the density of the troughs which, as a consequence, sink into the stable radiative zone below. The evolution of the tube is shown in Figure 1. Towards the top of the convection zone the background stratification becomes markedly more superadiabatic — the tube expands considerably and the field is weakened accordingly. The downflow of gas away from the crests reduces, but cannot prevent, this weakening.

Recently Chou & Fisher (1989) have published the results of a rather similar calculation — the key difference between their work and that of Moreno-Insertis (1986) lying in the treatment of the lower (non-buoyant) section of the tube. Whereas Moreno-Insertis essentially studied a tube with periodic undulations, Chou & Fisher have considered the rise of a loop of an otherwise straight tube. In the latter example the infinite reservoir of mass and flux in the straight section of the tube allows for a more effective drainage of the crests than in the periodic case, and consequently the velocity of rise is greater.

5.3 Magnetic Buoyancy as an Instability Mechanism

In this section we shall consider how magnetic buoyancy can act as an instability mechanism for any large-scale, diffuse field within the Sun. In §5.3.1 we review the linear theory of magnetic buoyancy instabilities of a static equilibrium under the assumption of ideal (diffusionless) MHD; in §§5.3.2, 5.3.3 we incorporate the important effects of finite diffusion and of rotation; the more recent nonlinear calculations are examined in §5.3.4. So as not to complicate matters unduly we shall not consider the effects of curved geometry — these may be found in the work of Acheson (1978). Thus, in essence, we shall be concentrating on instabilities that occur well outside the critical radius where a toroidal flux ring can be in equilibrium, the outwards buoyancy force and inwards curvature force being equal

(see Jensen 1955, Weiss 1964). This balance is attained at about one fifth of the solar radius and so a neglect of curvature effects is reasonable in describing instabilities in the convection zone or outer radiative zone. In all the subsequent discussion the magnetic field is assumed to be initially horizontal, in the y direction, and dependent only on the vertical coordinate z.

5.3.1 Ideal MHD

Magnetic buoyancy instability, in its simplest form, is best understood by neglecting all diffusive effects and considering modes that do not bend the magnetic field lines (so-called interchange modes). These may be dealt with by the following parcel argument (see, for example, Tayler 1973, Moffatt 1978, Acheson 1979b).

Imagine that a parcel of gas is raised, with no bending of the magnetic field lines, from a height z to a height $z+dz$. Suppose that the properties of the tube change from ϕ to $\phi+\delta\phi$, and let the local value of ϕ at height $z+dz$ be $\phi+d\phi$. Both the mass and magnetic flux of the parcel are conserved, giving the relation

$$\frac{\delta B}{B}=\frac{\delta\rho}{\rho}. \tag{5.4}$$

Assuming that the tube moves adiabatically, its pressure and density are related by

$$\frac{\delta p}{p}=\frac{\gamma\delta\rho}{\rho}. \tag{5.5}$$

Finally, we assume that the parcel moves sufficiently slowly that it is always in pressure equilibrium with its surroundings; thus

$$\delta p+B\delta B/\mu_0=dp+BdB/\mu_0. \tag{5.6}$$

The displaced parcel will be unstable and will continue to rise if $\delta\rho<d\rho$. Manipulation of (5.4) to (5.6) then gives the following criterion for instability:

$$\frac{-V_A^2}{\gamma H_p}\frac{d}{dz}\ln\left(\frac{B}{\rho}\right)>N^2, \tag{5.7}$$

where H_p ($\equiv p/g\rho$) is the pressure scale height, N is the Brunt-Väisälä or buoyancy frequency, and V_A is the Alfvén speed. Our main interest lies in cases where the atmosphere is convectively stable ($N^2>0$), any instability arising being due to the field stratification. It can be seen from (5.7) that if B/ρ decreases sufficiently rapidly with height then instability will ensue. As mentioned in the introduction, the instability is powered by the release of gravitational

potential energy made available by the field holding up more mass than would be possible in its absence. The local criterion (5.7) may be extended by the energy principle of Bernstein *et al.* (1958). This yields the result that instability of a general field distribution to interchange modes will occur if and only if (5.7) is satisfied anywhere in the fluid.

Things become considerably more complicated – even in the absence of diffusion – when perturbations that bend the field lines are allowed. Such modes were first considered by Newcomb (1961), who employed the energy principle, and by Parker (1966, 1967) who performed a normal-mode analysis for an isothermal equilibrium with constant Alfvén speed. For two-dimensional, undular perturbations confined to the yz-plane, Parker (1966) found that instability could occur provided that

$$\gamma<\frac{(\beta+1)^2}{\beta(\beta+3/2+2k_y^2H^2)}, \tag{5.8}$$

where β is the ratio of gas to magnetic pressure, k_y is the wavenumber of the horizontal perturbation and H is the scale height of the total pressure. Three-dimensional perturbations to the same atmosphere were found to be unstable when the following, less restrictive, condition is satisfied (Parker 1967):

$$\gamma<1+1/\beta, \tag{5.9}$$

where the limits of $k_y\to 0$, $k_x\to\infty$ have been taken. The major point to note from comparison of (5.7) with (5.8) or (5.9) is that instability to undular modes can occur for an atmosphere that is stable to interchange perturbations (for an isothermal atmosphere with constant Alfvén speed, B/ρ *in*creases with height and so (5.7) is never satisfied).

Hughes & Cattaneo (1987) extended the work of Newcomb (1961) to look in detail at the key differences between the interchange and undular modes and see precisely why the latter are favoured. For both types of mode the driving source is gravitational potential energy arising from the stratification of the field — it is the mechanism by which the energy is released that is the key difference. The density fluctuations necessary for liberating gravitational energy can only be created by doing a certain amount of work against pressure forces — the most vigorous instability will then be the one that releases the most potential energy for a given amount of work performed against pressure. Since interchange modes simply swap, but never bend, field lines, work must always be done against both gas *and*

magnetic pressure since, for such motions, the two are inseparable. By contrast, undular modes, for which the field lines are bent, have strong flows along the field and hence density fluctuations can be attained by doing work principally against the thermodynamic pressure. It is this ability of the undular modes to extract energy from the field stratification while escaping doing work against the magnetic pressure that causes them to be the more readily destabilised. (The above discussion has neglected the rôle of magnetic tension; although tension forces are stabilising, they can be made arbitrarily small by having sufficiently long wavelengths along the field.)

For two-dimensional undular modes (motions confined to the yz-plane) geometrical constraints generally make it impossible to extract the maximum benefit from the field stratification (see Hughes & Cattaneo 1987). However, with the extra freedom allowed for three-dimensional modes, the instability becomes maximally efficient when $k_y \to 0$, $k_x \to \infty$, with instability occurring if the following criterion is satisfied somewhere in the fluid (Newcomb 1961, Yu 1966, Thomas & Nye 1975, Hughes & Cattaneo 1987):

$$\frac{-1}{\gamma H_p}\frac{d}{dz}\ln B > \frac{N^2}{V_A^2}. \qquad (5.10)$$

(Gilman 1970 also considered these modes, but under the assumption of infinite, rather than zero, thermal conductivity. He found that instability could occur provided that B decreased upwards, all stabilising effects of the stratification being eroded by the rapid heat diffusion.) Comparison of inequalities (5.7) and (5.10) shows that whereas interchange modes require a decrease with height of B/ρ for instability, undular modes can be destabilised by just a decrease of B, a more readily satisfied condition. (Inequality (5.10) agrees, as it must, with (5.9) for the particular atmosphere studied by Parker.)

Finally in this subsection we shall consider the related instability of a discontinuous magnetic field. Although there are not so many subtleties involved as with a smoothly varying field, the nonlinear evolution in particular is of considerable interest. This is discussed in §5.3.4 — here we briefly consider the linear theory. As explained in the introduction, when the field is discontinuous there will typically also be a jump in the density. Hence, in the absence of inhomogeneities in the horizontal direction (and in contrast to the case of isolated flux tubes) there exist hydrostatic equilibria with magnetic gas supporting denser, non-magnetic gas. The source of potential energy available for driving an instability is very apparent in such

cases — at the magnetic interface the gas is not simply 'magnetically top-heavy' (B or B/ρ decreasing upwards) as in the earlier examples of this section, but *genuinely* top-heavy, with ρ increasing upwards. Consequently the instabilities of such equilibria are of the Rayleigh-Taylor type.

This problem was first tackled by Kruskal & Schwarzschild (1954) who considered the stability of a plasma supported above a vacuum by a uniform magnetic field. They found, by an essentially local analysis, that instability occurs provided that

$$2k_y^2 < |k_x|/H_p\,, \qquad (5.11)$$

where H_p is the pressure scale height of the plasma. The mechanism outlined above for a smoothly varying field, by which undular modes could be destabilised when interchanges are stable, is not relevant when the instability is driven only by a density jump at one height; consequently the most readily destabilised modes are interchanges ($k_y \equiv 0$), which do no work against magnetic tension. Not surprisingly, the instability criterion (5.11) is similar to that of the Rayleigh-Taylor instability of two different fluids, taking into account the effects of surface tension (see Chandrasekhar 1961); the key difference is that the magnetic field, in contrast to surface tension, imposes a preferred direction.

5.3.2 The Influence of Diffusion

As with many instabilities, incorporating the effects of finite diffusion leads to interesting new effects. The influence of diffusion is comparable for two- and three-dimensional modes and thus, to keep things reasonably straightforward, we shall confine our attention in this sub-section to interchange modes (motion confined to the xz-plane, all quantities invariant in y).

The gradients of temperature and magnetic field, together with thermal and magnetic diffusion, provide the ingredients for a double-diffusive system analogous in certain respects to thermohaline convection, possibly the most extensively studied of double-diffusive systems (see Turner 1973). Indeed, as shown by Spiegel & Weiss (1982) (see also Hughes & Proctor 1988), when magnetic buoyancy is incorporated into a modified Boussinesq approximation, the two cases may be formally identified via the following transformation:

$$R_A \to R_T + \frac{(\gamma-1)}{\gamma(1-\tau)}R_B, \qquad R_S \to \frac{(\gamma-\tau)}{\gamma(1-\tau)}R_B. \qquad (5.12)$$

Here R_A and R_S are the familiar thermal and solutal Rayleigh num-

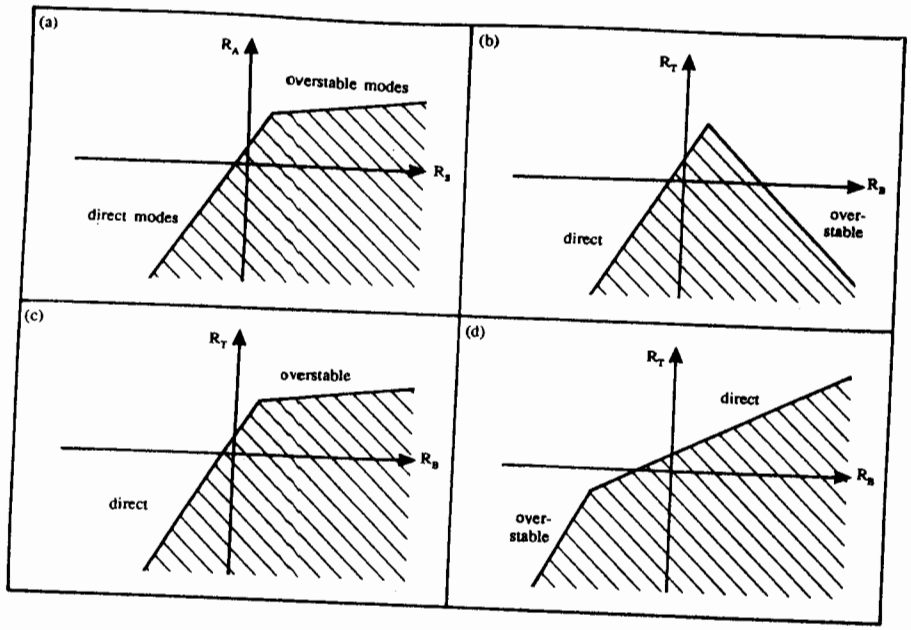


Figure 5.2. Sketches of regions of instability for thermohaline convection (a) and magnetic buoyancy (b, c and d) (from Hughes & Proctor 1988).

bers of thermohaline convection; increasing R_A or decreasing R_S is destabilising. R_T and R_B are the corresponding quantities in the magnetic problem, defined by

$$R_T = \frac{-d^4 N^2}{\kappa\nu}, \qquad R_B = \frac{gd^4}{\kappa\nu}\frac{B_0^2/\mu_0}{p_0}\frac{d}{dz}\ln\frac{B}{\rho}, \tag{5.13}$$

where d is the depth of the fluid layer, κ and ν are the thermal diffusivity and kinematic viscosity respectively, and p_0 and B_0 are typical values of the pressure and magnetic field within the layer. The atmosphere is convectively stable or unstable depending on whether R_T is negative or positive, and what we may term 'magnetically top-' or 'bottom-heavy' depending on whether R_B is negative or positive. The Schmidt number τ is the ratio of the diffusion of salt or magnetic field to that of heat.

If R_A and R_T were simply proportional to one another, as are R_S and R_B, then the two systems would behave virtually identically. However, the fact that the thermal Rayleigh number of thermohaline convection is mapped onto a combination of both Rayleigh numbers in the magnetic problem heralds the appearance of a new, somewhat counter-intuitive, instability. Figure 5.2a depicts the stability boundaries for thermohaline convection (with $\kappa > \kappa_s$) (Baines & Gill 1969,

Turner 1973). The physical mechanisms responsible for instability are readily understood in terms of simple parcel arguments. When the thermal gradient is stabilising and the salt gradient destabilising (the lower-left quadrant), then instability, when it occurs, does so via an exchange of stabilities (the growth rate passing through zero). Since the diffusion of heat occurs on a much faster time-scale than that of salt, a fluid parcel displaced upwards will find itself slightly cooler but considerably fresher than its surroundings; if it is less dense then its ascent will continue — a monotonically growing, or direct, instability. When, on the other hand, the thermal gradient is destabilising and the salt gradient stabilising (the upper-right quadrant), then instability can set in as growing oscillations (i.e. an overstability). This occurs by the following mechanism. A raised fluid parcel is hotter and saltier than its surroundings. If it is denser then it will fall, returning to its original level cooler but fresher than it was initially. If the net effect is such that the parcel returns with a greater density, then it will overshoot on the downward side. Repetition of this process leads to growing oscillations. As is to be expected, there is no instability in the lower-right quadrant where both gradients are stabilising. In the upper-left quadrant, where both gradients are destabilising, the tiny stable region is due solely to the damping effect of viscosity.

Figure 5.2b shows the corresponding stability boundaries for magnetic buoyancy, when $\tau < \gamma - 1 - \sigma$, where $\sigma = \nu/\kappa$ is the Prandtl number. The equation of the stability boundary for direct modes is identical for both systems; the instability criterion may be expressed as

$$\frac{-V_A^2}{H_p}\frac{d}{dz}\ln\frac{B}{\rho} > \nu\eta\frac{k^6}{k_x^2} + \frac{\eta}{\kappa}N^2, \tag{5.14}$$

where k_x is the horizontal wavenumber (recall that $k_y \equiv 0$), and k is the total wavenumber (Schubert 1968, Acheson 1979b). Inequality (5.14) is obviously just the diffusive modification to (5.7). In a convectively stable atmosphere, small values of η/κ are destabilising since the stabilising thermal gradient is eroded while the destabilising gradient of magnetic field is maintained.

The most unusual feature of Figure 5.2b is the appearance of instability in the lower-right quadrant. One's initial reaction is that this quadrant should be stable, as in the thermohaline case, since the gas is both subadiabatic ($R_T < 0$) and magnetically bottom-heavy ($R_B > 0$). However, as shown by Hughes (1985a), it is the

compressibility of the gas that permits a new instability not possible in thermohaline convection. When a flux tube is raised in an atmosphere where B/ρ increases with height it is squashed by the stronger external field and, if the squashing is sufficiently vigorous, it will be hotter than its surroundings. However, for $R_B > 0$ the tube will always be denser than its surroundings and will therefore fall. Instability is facilitated by thermal diffusion, which transmits heat away from the squashed (and heated) tube, causing it to return to its original position cooler, and hence denser, than it was initially. As explained above, this is precisely the recipe for overstability. The criterion for overstability is that the following inequality be satisfied (Hughes 1985a):

$$\frac{-V_A^2}{\gamma H_p}\left(\kappa(1-\gamma)+\nu+\eta\right)\frac{d}{dz}\ln\frac{B}{\rho} > (\kappa+\nu)(\kappa+\eta)(\nu+\eta)\frac{k^6}{k_x^2}+(\kappa+\nu)N^2 . \tag{5.15}$$

It can be seen that magnetic diffusion acts as a stabilising influence, eroding the squashing ability of the stronger external field and thereby reducing the compression of the tube and limiting its temperature rise. Indeed, for $\tau > \gamma - 1 - \sigma$, overstability in the lower-right quadrant ($R_T < 0$, $R_B > 0$) is completely suppressed.

Figure 5.2c depicts the stability boundaries when τ lies in the range $\gamma - 1 - \sigma < \tau < \gamma$, the stability boundaries still being given by (5.14) and (5.15). Overstability now sets in only when the gas is convectively unstable and magnetically bottom-heavy, the instability mechanism being analogous to that described earlier for thermohaline convection. If $\tau > \gamma$ (Figure 5.2d) then the rôles of the diffusivities are reversed and consequently the stability boundaries for the direct and overstable modes are exchanged.

5.3.3 The Influence of Rotation

Since the motivation for studying magnetic buoyancy instabilities is astrophysical, it is important to consider the influence of rotation. Here we shall restrict attention to the simplest case of uniform rotation about the x-axis, as a model of equatorial regions with predominantly azimuthal fields. Acheson (1978) has considered the more general case when rotation and gravity are not orthogonal, and where the rotation is not uniform; a whole host of different instabilities, such as the Goldreich-Schubert-Fricke instability and baroclinic instabilities, can then arise.

The effects of rotation were studied first by Gilman (1970) who

considered the influence of uniform rotation on an atmosphere with constant Alfvén speed, under the assumption of infinite thermal and electrical conductivities ($\kappa = \infty$, $\eta = 0$). Gilman found that sufficiently rapid rotation would cause all modes to become purely oscillatory. Several authors have extended Gilman's analysis to include more general effects. Roberts & Stewartson (1977) considered the same constant Alfvén speed atmosphere, but introduced the effects of finite diffusion of heat and magnetic field. After a very subtle scaling procedure with κ (suitably dimensionless) large, and η small, with $\eta\kappa \sim O(1)$, they performed a small perturbation analysis, essentially in η, to show that modes stable under the assumptions of Gilman ($\kappa = \infty$, $\eta = 0$) could now be unstable. Acheson & Gibbons (1978a) and Moffatt (1978) generalised Gilman's model to look at arbitrary field profiles, though restricting their attention to low-frequency modes obtained under the so-called magnetostrophic approximation. This consists in neglecting the inertial and viscous terms in the momentum equation, thus isolating modes characterised by a balance between Coriolis, Lorentz, buoyancy and pressure forces (see, for example, Acheson & Hide 1973). Acheson & Gibbons and Moffatt found that provided B/ρ decreases, locally, with height (which, as noted earlier, an atmosphere of constant Alfvén speed does not) then instability cannot be halted by rotation, however rapid.

With the inclusion of rotation, the dispersion relation governing the growth rates of three-dimensional diffusive modes becomes rather complicated (fifth order with complex coefficients), and it is not a straightforward matter to write down general criteria for instability. Under the assumptions of rapid rotation and rapid heat diffusion, and concentrating on the low-frequency magnetostrophic modes, Acheson (1978) derived the following condition for instability of modes driven by a top-heavy magnetic field:

$$\frac{-1}{H_p}\frac{d}{dz}\ln\frac{B}{\rho} > \frac{k_y^2k^2}{k_x^2}+\frac{4\Omega^2}{V_A^4}\eta^2\frac{k^4}{k_y^2}+\left(\frac{\gamma-1}{2\gamma H_p}\right)^2+\frac{\eta}{\kappa}N^2V_A^2 - \frac{\left((\gamma+1)/2\gamma H_p\right)^2}{\left(1+8\eta\kappa(\Omega^2/N^2V_A^2)(k^4/k_y^2)\right)^2}, \tag{5.16}$$

with instability setting in as travelling waves. Acheson used this relation to tie together nicely the results of Gilman (1970), Roberts & Stewartson (1977) and Acheson & Gibbons (1978a).

Within the magnetostrophic approximation, instability may also occur when B/ρ increases with height (Hughes, 1985b). Although

the driving force is as discussed in §5.3.2, with the field acting as a squashing, and therefore heating, agent, it should be noted that the modes are now waves influenced strongly by rotation and that they have a very different character to their counterparts in the non-rotating case.

From inequality (5.16) it can also be seen that instability may be facilitated by an increase in the stable stratification (i.e. an increase in N^2). In §5.3.2 we saw that in the absence of rotation, diffusion was absolutely essential for the occurrence of overstable oscillations; in a rotating system, although instability may (and typically does) occur in an oscillatory manner, even in ideal MHD, the rôle of diffusion is still important in determining the density changes sustained during an oscillation. Thus, crudely speaking, the interesting phenomenon of destabilisation by increasing the stable stratification may be explained as the oscillations being pushed towards a frequency where diffusion is more effective (or possibly less destructive) for driving overstability. More general instabilities of this nature have been considered by Acheson (1980); related instabilities in other fluid dynamical systems have also been found (see, for example, Roberts 1978, Masuda 1978).

Although the magnetostrophic approximation leads to a valuable simplification of the governing equations, certain modes, which may be of importance, are of course omitted. Interchange modes, which have no magnetic tension, are excluded, as are high-frequency modes, for which $|\partial_t \mathbf{u}|$ is comparable to $|\Omega \times \mathbf{u}|$. In the absence of viscosity, rotation has a dramatic stabilising effect on direct interchange instabilities, caused by the angular momentum constraint imposed on such motions (this is broken in *non*-axisymmetric motions by the azimuthal component of the Lorentz force). Viscosity relaxes this constraint, an exchange of stabilities setting in when

$$\frac{-V_A^2}{H_p}\frac{d}{dz}\ln\frac{B}{\rho} > 4\Omega^2\eta/\nu + N^2\eta/\kappa + \eta\nu\frac{k^6}{k_x^2}. \qquad (5.17)$$

Direct modes are the dominant form of axisymmetric instability for top-heavy gradients for very small values of τ. When $\tau \gtrsim \sigma$ (for small σ) the preferred mode at the onset of instability becomes oscillatory. The régime of $\sigma \ll \tau \ll 1$ has been studied in detail by Acheson & Gibbons (1978b), who explain how the instability takes place in an oscillatory fashion, to overcome the rotational constraint, but with a short (though not too short!) wavelength which diffuses heat rapidly,

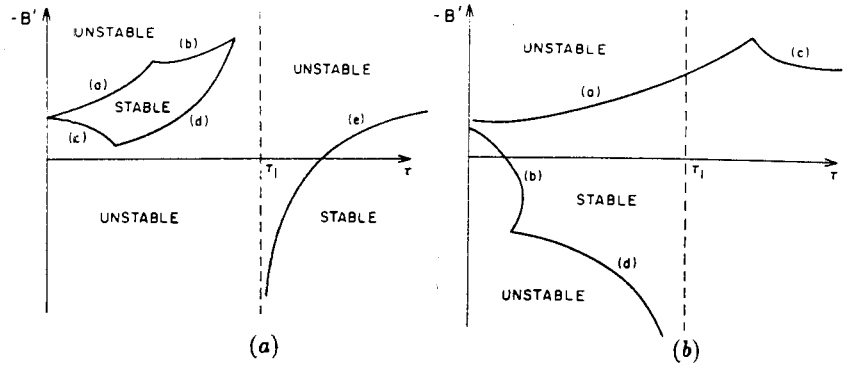


Figure 5.3. Sketches of the overall stability boundaries for rotating magnetic buoyancy instabilities. In (*a*) the Prandtl number σ is small and the stable stratification weak. Line (a) denotes the boundary of the top-heavy (magnetostrophic) mode, line (b) that of the triply-diffusive interchange instability, and lines (c), (d) and (e) the stability boundaries of the high-frequency modes. An increase in σ or N^2 transforms the stability boundary into that shown in (*b*). Line (a) is as before, (b) is the boundary of the low-frequency, bottom-heavy mode, (c) and (d) the high-frequency boundaries (from Hughes 1985b).

thereby annulling the stable stratification, but not allowing excessive viscous damping (recall that $\nu \ll \kappa$).

The interchange instabilities driven by bottom-heavy field gradients are still present in a rotating system. Indeed, initially increasing the rotation rate is destabilising, by a similar mechanism to that outlined above involving the convective stratification; however, sufficiently rapid rotation is stabilising, as one might imagine.

The high-frequency undular modes are oscillatory in character, but not dramatically different to their non-undular (interchange) counterparts (i.e. in the limit $k_y \to 0$ such modes tend smoothly to interchanges with $k_y \equiv 0$). Instability is prevalent for both top- and bottom-heavy magnetic field gradients when the viscous damping and the stable stratification are weak; however, since the growth rate of these modes is small ($O(\Omega^{-1})$ for rapid rotation) they can easily be killed off by moderate stabilising influences. Figure 3 is a sketch of the overall stability boundary (i.e. the boundary between stability to *all* modes and instability to *any* mode), with the critical magnetic field gradient plotted against τ $(= \eta/\kappa)$.

Up to now we have considered the case of rapid rotation, for which there is a well-defined separation into low- and high-frequency modes. For slower rotation rates this distinction becomes blurred and things

become considerably more complicated with rather bizarre stability boundaries — the full gory details may be found in Hughes (1985b).

5.3.4 The Nonlinear Evolution of Magnetic Buoyancy Instabilities

The considerable increase in computing power over the past decade or so has allowed the detailed study of many nonlinear hydrodynamical and hydromagnetic processes, including the nonlinear evolution of magnetic buoyancy instabilities. Cattaneo & Hughes (1987, 1988) addressed the issue of how a strong toroidal field in the overshoot zone could break up and escape. They modelled, in Cartesian geometry, the axisymmetric instabilities of a toroidal field at the equator; the initial state consisted of a slab of uniform magnetic field (in the y-direction) embedded in a convectively stable piecewise-polytropic atmosphere, the sole source of instability being the density jump at the upper magnetic interface. Only interchange modes (motions confined to the xz-plane) were considered but, as noted earlier, this is a reasonable assumption for Rayleigh-Taylor instabilities. The nonlinear evolution of the magnetic field, following a small random perturbation of the initial state, is shown in Figure 4. A strong shear develops between the rising magnetic field and the falling non-magnetic fluid, causing the onset of Kelvin-Helmholtz instabilities and the subsequent wrapping up of the gas into regions of strong vorticity, located in the wings of the mushrooms of Figure 4b. Once most of the available gravitational potential energy has been released, the vorticity distribution plays a key rôle in the subsequent evolution of the magnetic layer. Adjacent vortices on neighbouring mushrooms interact so as to move the vortex pair downwards, and indeed these interactions can be of sufficient strength that regions of strong field (which are lighter than their surroundings) are pushed down. This rise and subsequent fall of the field is illustrated clearly by the pair of mushrooms to the left of centre in Figures 5.4b,c,d. The final stage (Figure 5.4f) shows a well-dispersed field, the complex motions arising from the interacting vortices having spread the field over the entire domain in a time short compared to the magnetic diffusion time.

Cattaneo, Chiueh & Hughes (1990a,b) have extended these calculations to investigate whether the addition of a weak poloidal field alters the character of the instability. The initial field was changed from being uni-directional to having the form $\mathbf{B} = (B_x(z), \sqrt{(1 - B_x^2)}, 0)$. Attention was again restricted to y-independent (axisymmetric) motions, although the velocity and magnetic fields now have components in all three directions. A key feature of this field configuration is the possible existence of a so-called *resonant surface* on which the poloidal field vanishes — as we shall see, the location of this surface plays a crucial rôle in the evolution of the instability. (Although the existence of a resonant surface in the Cartesian model might seem rather arbitrary, since a rotation of the coordinates about the z-axis swaps the 'poloidal' and 'toroidal' components, the uniqueness of the resonant surface is ensured by noting that the Cartesian model is constructed as a local representation of the system in spherical geometry where the poloidal and toroidal components *are* uniquely specified.)

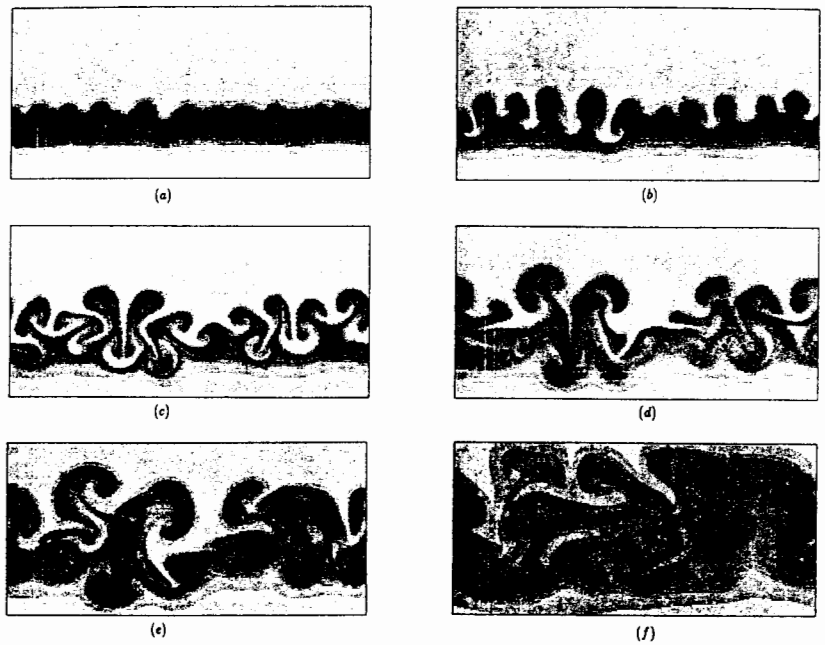
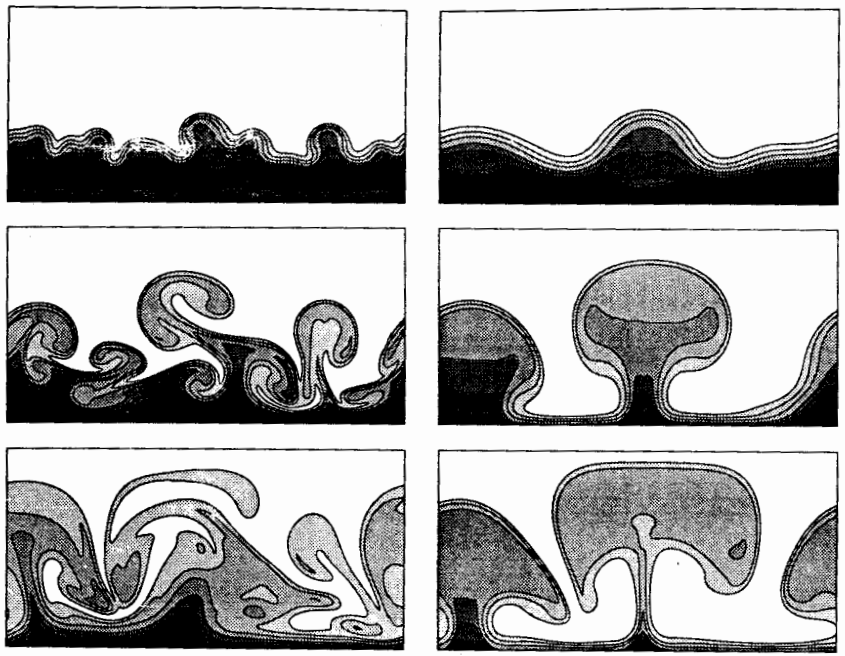


Figure 5.4. Time evolution of an unstable magnetic layer. The field lines remain straight, into the plane of the page, for all times. Darker regions correspond to stronger field (from Cattaneo & Hughes 1987).

Including a poloidal component of the field brings into play mag-



(*a*) (*b*)

Figure 5.5. Time evolution of a predominantly, though not exclusively, toroidal field. Darker regions correspond to stronger toroidal field. The equilibrium in (*a*) has the poloidal field vanishing near the magnetic interface and increasing with depth; in (*b*) it vanishes near the lower boundary and increases with height. The motions are independent of the azimuthal direction (into the page).

netic tension and, in contrast to the pure interchange modes described above, work must now be done in untwisting the magnetic field lines. This is minimised in the neighbourhood of the resonant surface, where the field lines locally are untwisted. Figure 5.5 shows the evolution of the instability from two equilibria, both of which have a poloidal field varying quadratically with height, but which differ in the location of the resonant layer. The equilibrium of Figure 5.5a has the resonant layer just beneath the magnetic interface. The instability is confined to the upper regions of the field, where the poloidal component is weak, and is not dissimilar to the evolution of

the purely toroidal field of Figure 5.4. Penetration into the lower part of the layer is hindered by the increasing poloidal field and thus part of the layer avoids disruption. The equilibrium of Figure 5.5b, on the other hand, has a resonant layer close to the bottom boundary, and it can be seen that the evolution is markedly different. The poloidal field is strongest close to the interface, the associated magnetic tension leading to a preferred mode of long horizontal scale. Untwisting the magnetic field becomes easier with depth, as the poloidal field weakens, and so the unstable mode extends from the magnetic interface all the way down to the resonant surface. In consequence, and in contrast to the case of Figure 5.5a, little remains of the magnetic layer after the instability has taken place. In addition to the considerable difference in scale, the ejected flux of the two cases differ in one other very important respect. When the resonant layer is close to the interface (Figure 5.5a) the ejected field is predominantly toroidal and its behaviour is thus like that of the interchange modes of a purely toroidal field — Kelvin-Helmholtz instabilities set in, wrapping the gas into pockets of strong vorticity that interact and subsequently lead to a very effective mixing of magnetic flux. However, when the resonant layer is deep-seated, inspection of Figure 5.5b shows that the strongest poloidal field in the ejected magnetic fragments lies at their outer edges. The wrapping up of the gas into regions of strong vorticity is thereby resisted most effectively; the fragments are thus not distorted and rise unscathed.

The nonlinear evolution of magnetic buoyancy instabilities in the Sun has also been considered by Shibata *et al.* (1989, 1990), developing earlier work of Matsumoto *et al.* (1988) concerning the magnetic field in an accretion disc. Shibata *et al.* (1989) modelled the emergence of the field from the chromosphere into the corona by considering an equilibrium in which a slab of horizontal magnetic field of constant β is embedded in an isothermal atmosphere (the chromosphere), above which lies another, hotter, isothermal atmosphere (the corona). There are no discontinuities as such, the temperature and field varying smoothly (though very rapidly in places) in such a manner so as just to avoid genuine top-heaviness in the gas. The equilibrium is perturbed by two-dimensional undular perturbations; the effects of diffusion are neglected, and a value of γ close to unity is chosen so as to facilitate instability (see (5.8)). The field rises as a loop through the chromosphere before being decelerated in the hotter corona. Shibata *et al.* (1990) perform a similar calculation, though now adding a convection zone underneath to house the initial field.

In this model the field is discontinuous, leading to a destabilising jump in the density at the upper magnetic interface. Thus, besides instability arising from the field stratification there is also the possibility of Rayleigh-Taylor modes although, somewhat surprisingly, Shibata *et al.* make no mention of this. The general evolution of the magnetic field is rather similar to that of their earlier calculation.

5.4 Discussion

Having considered the physical ideas behind the two mechanisms of magnetic buoyancy we are now in a position to discuss their effects on the solar field. Although this review has concentrated on the rise of isolated tubes and the break-up of large-scale field distributions, these are of course just part of a larger sequence of processes involving the generation of the solar field and its subsequent escape from the interior. It is widely believed that the Sun's field is maintained by some sort of $\alpha\omega$ dynamo, with a regeneration time of a few years, the most obvious location for such a dynamo being the convection zone — a region of helical motions and differential rotation. However, as we shall discuss below, there are certain difficulties with this idea and indeed there is as yet no entirely satisfactory description of the solar dynamo. (It should though also be noted that alternative, oscillator theories of the cyclical nature of the solar field are, at least in their current state, considerably less plausible — see Cowling 1981 for a critical review of the two types of theory.) The nature of any dynamo process is tied in closely with the form of the solar field at depth — the results of §§5.2 and 5.3 illustrate the very different behaviour of isolated tubes as opposed to diffuse distributions and it is therefore an important question as to which of these field distributions occurs within the Sun. We know that at the surface the field tends to be in discrete bundles, and the appearance of large active regions certainly suggests the emergence of isolated tubes from depth. Whether the solar field is always in such a tubular state however, or if instead flux tubes are formed from a large-scale distribution, remains a matter of debate.

There are two principal difficulties with the assumption that the field in the Sun exists always in the form of isolated tubes of flux. The first is that flux-tube dynamos are rather hard to construct (see, for example, the review by Schüssler 1983); the second is that, as we have seen in §5.2, isolated flux tubes are inherently buoyant and are hard to trap within the convection zone for times comparable to the

regeneration timescale of the field. Several remedies of the latter problem have been suggested.

The simplest is to suppose that the flux tubes are confined to the convectively stable radiative zone. The elementary arguments expounded in the introduction show that tubes in thermal equilibrium with their surroundings are buoyant, regardless of the stratification of the external medium. However, whereas adiabatically expanding tubes are yet more buoyant (than ones always in thermal equilibrium) when the external atmosphere is superadiabatic (see §5.3), in a stable atmosphere a tube initially in thermal equilibrium but rising adiabatically will eventually come into mechanical equilibrium (i.e. with $\rho_i = \rho_e$) — the rate of heat flow between tube and surroundings is therefore a crucial factor in determining the rate of rise. Parker (1975) considered the quasi-static rise of tubes with heat flow governed by Newton's law of cooling — he found that the speed of rise was extremely slow (only $\sim$ 6 cm/s for a tube of radius 300km and field strength 10^7G) and that the tubes would be effectively anchored for times long compared to the solar cycle period (though not compared to the age of the Sun). Thus the surface toroidal field, which is observed to change sign approximately every 11 years, cannot be a result of the escape of flux tubes from the radiative interior.

Since it seems that the radiative zone hangs on to flux tubes for times long compared to the solar cycle, whereas the convection zone releases them rather readily, it is natural to enquire as to whether the overshoot zone, the convectively stable transition region between the two, is a suitable storage location. Van Ballegooijen (1982) concludes that even tubes initially neutrally buoyant will leave the overshoot zone within a year as a result of buoyancy accrued from heat inflow — this presumably is an overestimate since it is unlikely that tubes will find themselves sufficiently cool as to start off in mechanical equilibrium. He goes on to suggest that flux tubes may however be confined to the overshoot region for longer times by an inward Coriolis force arising from advection by an equatorward meridional circulation. Van Ballegooijen & Choudhuri (1988) have considered the stability of tubes held in equilibrium by a combination of buoyancy, tension, Coriolis and drag forces and find that they can be stable in the lower part of the convection zone. Parker (1987 and subsequent papers) has advanced the alternative suggestion that the buoyancy of flux tubes might be overcome, even in the lower regions of the convection zone, by what he terms 'thermal shadows'. The idea is that large flux tubes can suppress the convective heat flow

from below and that the resulting cool shadow above will overcome the tube's buoyancy. Magnetic field is however eventually brought to the surface when the accumulation of heat underneath the tube is sufficient to cause an eruption of fluid through the field. Parker proposes that this phenomenon will be of an oscillatory nature —after such an eruption the field will close and the whole process of heat accumulation will begin afresh.

As an alternative to the notion of the field always being in the form of discrete tubes, it may exist in a more homogeneous or diffuse state, with the isolated tubes that we see at the surface resulting from some secondary process. Again there are problems in confining such a field to the convective zone. Although recent numerical simulations have demonstrated the plausibility of self-consistent – rather than just kinematic – convection-zone dynamos, the models differ in key respects to the observed solar field. In particular, the cycle time is rather short and magnetic features migrate poleward, rather than towards the equator as on the Sun. (A clear account of recent work in solar dynamo theory, together with appropriate references, can be found in the review of DeLuca & Gilman 1990.) The numerical models indicate that the helicity in the overshoot zone is of opposite sign to that in the convection zone above and, at least on kinematic grounds, this suggests that a dynamo located in the overshoot region should display the correct (i.e. solar) propagation direction — the results of Glatzmaier (1985) are encouraging, though not conclusive, on this score.

Besides problems arising from numerical simulations of the dynamo process, there are other important, and indeed older, objections to there being a diffuse field throughout the convection zone, these relating to the interaction of magnetic fields and convection. Galloway & Weiss (1981) point out that flux will tend to be expelled from convective cells and form concentrated magnetic ropes at the boundaries — these will be buoyant and, unless trapped in a region of strong downflow, may be expected to ascend through the convection zone. Galloway & Weiss (see also Spiegel & Weiss 1980) suggest that in order to supply the flux observed emerging at the surface the field must be confined to the overshoot region, which would also be the site of dynamo action. It is difficult to be more precise on this matter since so little is known about what is going on in the bulk of the convection zone, beneath the surface phenomena of granulation and supergranulation. For example, is the influence of rotation sufficiently strong that the Taylor-Proudman constraint holds and the

convection is in cells elongated parallel to the rotation axis (banana- or cartridge-belt cells)? If so, what is the lifetime of such cells? Or might the convection be somewhat less regular, as in some of the Spacelab experiments of convection in a rotating spherical shell (see Hart, Glatzmaier & Toomre 1986)?

Now if the field *is* in a diffuse state, in the convectively stable overshoot zone, it is necessary to explain how the tubes that we see at the surface are formed. One possibility is that every now and then a particularly exuberant convective motion sweeps up a stitch of magnetic flux. A more systematic means of flux loss would be due to instabilities driven by magnetic buoyancy, of the form described in §3. There exists the possibility of instability resulting from a smooth gradient in the field or from a discontinuity at the upper boundary — or indeed a combination of the two. Schmitt & Rosner (1983) and Hughes (1984, 1985b) have considered the types of field gradient instability that might occur, though from slightly differing viewpoints. Schmitt & Rosner considered a fixed value of the magnetic scale height ($H_B = H_\rho/5$) and looked at the variation of growth rates due to changes in the diffusivities and in the scale of the motions. They concluded that both low- and high-frequency modes could be unstable, with comparable growth rates. Hughes adopted a complementary approach of assuming the diffusivities and subadiabatic stratification to be known and then calculating which field gradients could be unstable. If the diffusivities take their molecular values, with $\nu \ll \eta \ll \kappa$, then *all* field gradients are found to be locally unstable — the most rapidly growing is the low-frequency mode driven by a field decreasing with height. If on the other hand the diffusivities take some sort of eddy values with $\nu \sim \eta \sim \kappa$ then all instabilities due to bottom-heavy fields are killed off. Instability then occurs only for rather top-heavy fields ($H_B = H_\rho/400$).

Of course, although a linear stability analysis gives a useful guide to the initial tendency to instability, it can shed no light on whether flux tubes can emerge from a destabilised layer. As described in §5.3.4, Cattaneo & Hughes (1987, 1988) and Cattaneo *et al.* (1990a,b) have addressed this issue by studying the nonlinear development of Rayleigh-Taylor instabilities — although a strict discontinuity in the field is an idealisation, similar behaviour is expected in any thin magnetic layer where the field decreases upwards rapidly at some point. If the field is purely toroidal then fragments of strong field are ejected (as the mushrooms of Figure 5.4) — however, subsequent vortex interactions lead to a general diffusion of flux and also to the

downward transport of the pockets of strong field. The inclusion of a weak poloidal ingredient – as of course there must be in the Sun – makes a major difference (Figure 5.5). As explained earlier, the position of the resonant surface determines whether the field emerges as small-scale, readily distorted blobs or as larger structures which, as they rise, maintain their coherence due to their helical field distribution. As pointed out by Cattaneo *et al.* (1990b) small variations in the distribution of poloidal field during the solar cycle might account for the observed modulations in scale, morphology and amplitude of emerging flux (Golub *et al.* 1981).

Finally it only remains to discuss briefly some of the interesting outstanding problems that could be tackled to improve our understanding of the behaviour of the large-scale solar magnetic field. My own personal feeling is that the field does not exist always in discrete tubes since, besides the problems of flux-tube dynamos, a conspiracy of forces is required to hold them down. Instead it seems more likely that a dynamo produces a diffuse field, with isolated tubes arising via an instability mechanism as described above. A key problem, regardless of the precise nature of how flux tubes come about, concerns their interaction with the convection as they rise to the surface. It can be argued, with some justification, that the ordered emergence of tubes shows that they have not been severely distorted by the convection *en route*; if this is so, then it is of interest to know just how the convection adjusts to let the tubes through. Also, even if in the lower part of the convection zone the tubes are sufficiently strong to resist the convection, as they rise they expand considerably (Moreno-Insertis 1986) and some sort of concentration by convection seems desirable. There is the further question of what happens when the thin flux tube assumption is abandoned. Schüssler (1979) has shown how a rising tube can fragment; the nonlinear results of §5.3.4 suggest that the vorticity distribution in the tube, as well as its helical field structure, may be of importance.

On the issue of the break-up of a magnetic layer, there are obvious 3-dimensional extensions to the calculations of §5.3.4 — though only for those both brave and computationally wealthy. Possibly a more fundamental issue could be addressed by incorporating an overlying superadiabatic atmosphere and thereby including some of the effects of convective overshoot into the model.

Acknowledgement

I am very grateful to Professor N.O. Weiss for helpful discussions concerning this review.

References

Acheson, D.J. 1978 *Philos. Trans. R. Soc. London* **A289**, 459-500. Acheson, D.J. 1979a *Nature* **277**, 41-2.

Acheson, D.J. 1979b *Solar Physics* **62**, 23-50.

Acheson, D.J. 1980 *J. Fluid Mech.* **96**, 723-33.

Acheson, D.J. & Hide, R. 1973 *Rep. Prog. Phys.* **36**, 159-221.

Acheson, D.J. & Gibbons 1978a *J. Fluid Mech.* **85**, 743-57.

Acheson, D.J. & Gibbons 1978b Appendix to Acheson (1978).

Baines, P.G. & Gill, A.E. 1969 *J. Fluid Mech.* **37**, 51-80.

Bernstein, I.B., Frieman, E.A., Kruskal, M.D. & Kulsrud, R.M. 1958 *Proc. R. Soc. London* **244**, 17-40.

Cattaneo, F. & Hughes, D.W. 1987 In *Theoretical Problems in High Resolution Solar Physics II* (eds. G. Athay and D.S. Spicer) NASA CP-2483, pp101-4.

Cattaneo, F. & Hughes, D.W. 1988 *J. Fluid Mech.* **196**, 323-44.

Cattaneo, F., Chiueh, T. & Hughes, D.W. 1990a In *Topological Fluid Mechanics* (eds. H.K. Moffatt and A. Tsinober) pp679-88 C.U.P.

Cattaneo, F., Chiueh, T. & Hughes, D.W. 1990b *J. Fluid Mech.* (in press).

Chandrasekhar, S. 1961 *Hydrodynamic and Hydromagnetic Stability.* Clarendon.

Chou, D.Y. & Fisher, G.H. 1989 *Astrophys. J.* **341**, 533-48.

Choudhuri, A.R. & Gilman, P.A. 1987 *Astrophys. J.* **316**, 788-800.

Cowling, T.G. 1981 *Ann. Rev. Astron. Astrophys.* **19**, 115-35.

DeLuca, E.E. & Gilman, P.A. 1990 In *The Solar Atmosphere and Interior* (eds. A.N. Cox, W.C. Livingston and M.S. Mathews) Univ. of Arizona Press.

Galloway, D.J. & Weiss, N.O. 1981 *Astrophys. J.* **243**, 945-53.

Gilman, P.A. 1970 *Astrophys. J.* **162**, 1019-29.

Golub, L., Rosner, R. Vaiana, G.S. & Weiss, N.O. 1981 *Astrophys. J.* **243**, 309-16.

Glatzmaier, G.A. 1985 *Geophys. Astrophys. Fluid Dyn.* **31**, 137-50.

Hart, J.E., Glatzmaier, G.A. & Toomre, J. 1986 *J. Fluid Mech.* **173**, 519-44.

Hughes, D.W. 1984 In *The Hydromagnetics of the Sun. Proceedings 4th European Meeting on Solar Physics.* ESA SP-220 pp55-58.

Hughes, D.W. 1985a *Geophys. Astrophys. Fluid Dyn.* **32**, 273-316.

Hughes, D.W. 1985b *Geophys. Astrophys. Fluid Dyn.* **34**, 99-142.

Hughes, D.W. & Cattaneo, F. 1987 *Geophys. Astrophys. Fluid Dyn.* **39**, 65-81.

Hughes, D.W. & Proctor, M.R.E. 1988 *Ann. Rev. Fluid Mech.* **20**, 187-223.

Jensen, E. 1955 *Ann. Astrophys.* **18**, 127-40.

Kruskal, M. & Schwarzschild, M. 1954 *Proc. R. Soc. London* **A223**, 348-60.

Masuda, A. 1978 *J. Ocean. Soc. Japan* **34**, 8-16.

Matsumoto, R., Horiuchi, T., Shibata, K. & Hanawa, T. 1988 *Publ. Astron. Soc. Japan* **40**, 171-95.

Moffatt, H.K. 1978 *Magnetic Field Generation in Electrically Conducting Fluids* C.U.P.

Moreno-Insertis, F. 1983 *Astron. Astrophys.* **122**, 241-50.

Moreno-Insertis, F. 1986 *Astron. Astrophys.* **166**, 291-305.
Newcomb, W.A. 1961 *Phys. Fluids* **4**, 391-6.
Parker, E.N. 1955 *Astrophys. J.* **121**, 491-507. Parker, E.N. 1966 *Astrophys. J.* **145**, 811-33.
Parker, E.N. 1967 *Astrophys. J.* **149**, 535-52.
Parker, E.N. 1969 *Space Science Rev.* **9**, 651-712.
Parker, E.N. 1975 *Astrophys. J.* **198**, 205-9.
Parker, E.N. 1987 *Astrophys. J.* **312**, 868-79.
Roberts, P.H. & Stewartson, K. 1977 *Astron. Nachr.* **298**, 311-8.
Roberts, P.H. 1978 In *Rotating Fluids in Geophysics* (eds. P.H. Roberts and A.M. Soward) pp420-36 Academic Press.
Schmitt, J.H.M.M. & Rosner, R. 1983 *Astrophys. J.* **265**, 901-24.
Schubert, G. 1968 *Astrophys. J.* **151**, 1099-1110.
Schüssler, M 1977 *Astron. Astrophys.* **56**, 439-42.
Schüssler, M 1979 *Astron. Astrophys.* **71**, 79-91.
Schüssler, M 1983 In *Solar and Stellar Magnetic Fields. IAU Symp. 102* (ed. J.O. Stenflo) pp213-36 Reidel.
Shibata, K., Tajima, T., Matsumoto, R., Horiuchi, T., Hanawa, T., Rosner, R. & Uchida, Y. 1989 *Astrophys. J.* **338**, 471-92.
Shibata, K., Nozawa, S., Matsumoto, R., Sterling, A.C. & Tajima, T. 1990 *Astrophys. J.* **351**, L25-8.
Spiegel, E.A. & Weiss, N.O. 1980 *Nature* **287**, 616-7.
Spiegel, E.A. & Weiss, N.O. 1982 *Geophys. Astrophys. Fluid Dyn.* **22**, 219-34.
Spruit, H.C. & van Ballegooijen, A.A. 1982 *Astron. Astrophys.* **106**, 58-66.
Stella, L. & Rosner, R. 1984 *Astrophys. J.* **277**, 312-21.
Tayler, R.J. 1973 *Mon. Not. R. Astron. Soc.* **161**, 365-80.
Thomas, J.H. & Nye, A.H. 1975 *Phys. Fluids* **18**, 490-1.
Turner, J.S. 1973 *Buoyancy Effects in Fluids.* C.U.P.
Unno & Ribes 1976 *Astrophys. J.* **208**, 222-3.
van Ballegooijen, A.A. 1982 *Astron. Astrophys.* **113**, 99-112.
van Ballegooijen, A.A. & Choudhuri, A.R. 1988 *Astrophys. J.* **333** 965-77.
Weiss, N.O. 1964 *Mon. Not. R. Astr. Soc.* **128**, 225-35.
Yu, C.P. 1966 *Phys. Fluids* **9**, 412-4.

6

Magnetohydrodynamic Waves in the Sun

B. ROBERTS

The University, St Andrews, KY16 9SS, Scotland

6.1 Introduction

The dynamics of the solar atmosphere is complicated by the fact that not only is it strongly stratified in both gas density and temperature, but it is also permeated by a non-uniform magnetic field. The solar atmosphere is the example, *par excellence*, of a plasma that is both structured and stratified. Structuring arises because of the powerful ordering of the magnetic field-lines that emerge through the photosphere into concentrated flux tubes, an effect due to the strong convective processes operating in this region of the Sun. Stratification is important in the photospheric layers too, for in this region of the Sun pressure and density vary most rapidly with height. Elsewhere, either deeper in the interior or higher in the corona, the effects of stratification are somewhat less marked.

What is the nature of wave motions in such a medium? The complexity of an atmosphere that is both magnetically structured and gravitationally stratified forbids any simple answer to such a question, and one has to examine separately a number of differing aspects of our problem in order to shed light on the overall nature of the Sun as an oscillator.

There is no surprise that the Sun is able to sustain oscillations of various kinds. After all, the Sun is a compressible plasma and so able to support sound waves. The presence of magnetic field means that the solar atmosphere is a particularly elastic medium, with an elasticity that is additional to the natural compressibility that a gas possesses. Magnetic fields impart magnetic pressure and tension forces which help to sustain such wave motions. Thus, that wave motions occur is no surprise. What is less clear, however, is what sustains wave motions: what effects operate so as to disturb

the equilibrium state of the magnetized gas, making oscillations the natural form?

The structured nature of the solar magnetic field means that magnetism is of greater importance in some regions of the Sun than in others. A simple guide to the relative importance of magnetic effects is provided by the plasma beta, β, defined by

$$\beta = \frac{gas\ pressure}{magnetic\ pressure}, \tag{6.1}$$

the magnetic pressure in a field of strength B_0 tesla being $B_0^2/2\mu$ newtons per square metre, for magnetic permeability μ ($= 4\pi \times 10^{-7}$ in mks units). (In cgs units, a field of B_0 gauss has an associated magnetic pressure of $B_0^2/8\pi$ dynes cm^{-2}.) The solar interior is generally considered to be a region of high β, because the gas pressure there is so great, though it should be noted that there is considerable uncertainty about the field strength in the Sun's interior. So magnetic effects are generally not expected to be large in the layers of the Sun below the photosphere, though it is here that the magnetic dynamo (believed responsible for generating the Sun's field) must reside. By contrast, the rapid decrease in gas pressure in the upper layers of the solar atmosphere suggests that β is of order unity or smaller there, and so magnetic effects are then likely to be particularly significant. In the layers in between, covering the photosphere and low chromosphere, β is roughly of order unity wherever there are magnetic fields.

What observational evidence is there for waves in the Sun? The first evidence for an oscillation in the Sun came from Leighton, Noyes and Simon (1962), who detected packets of oscillation in the photosphere with indications of a five-minute periodicity. The nature of these oscillations was not fully understood at the time and had to await observational developments by Deubner (1975) and theoretical descriptions by Ulrich (1970) and Leibacher and Stein (1971). With these developments it soon after became clear that the five-minute modes were in fact global acoustic modes of oscillation of the Sun as a whole, the so-called p-modes. Thus we have the remarkable fact that the Sun is vibrating like a drum, save that the interior as well as the solar surface partakes in the modes of oscillation. Effectively, the rising sound speed in the solar interior forms a cavity within which sound waves (p-modes) are confined; the cavity extends from roughly the solar surface to a depth which depends upon the wavelength of the mode.

Recent detailed studies of the Sun's modes of oscillation are exhibiting the internal structure of the Sun in a way that no other means can: for the first time, we are able to 'look' below the solar surface. Much as the seismologist has learnt about the Earth's interior from a study of its modes of oscillation excited by earthquakes, the new subject of *helioseismology* (see reviews by Deubner and Gough 1984; Christensen-Dalsgaard, Gough and Toomre 1985; Leibacher *et al.* 1985) promises much in our attempts to understand the solar interior.

The global modes of oscillation of the Sun are likely to be influenced by the magnetism of the chromosphere (see Campbell and Roberts 1989; Evans and Roberts 1990b), but such effects are relatively small as regards the intrinsic character of the modes. One has to look at regions of the Sun where magnetism is central to see oscillations that are strongly influenced by magnetic fields. Sunspots, the seats of 0.3 tesla (3 kilogauss) magnetic fields, provide such locations.

Oscillations in sunspots were first detected by Beckers and Tallant (1969) and subsequent studies have revealed an ever greater degree of complexity (see the review by Moore and Rabin 1985). Sunspots exhibit both five- and three-minute oscillations in their umbrae, with the three-minute modes being the stronger. Sunspot penumbrae also exhibit oscillations, with periods of around four to five minutes or larger (see Lites 1988). Sometimes coherent wave-fronts, apparently emanating from the umbra, are seen to propagate out across the penumbra with speeds of 20-35 km s^{-1}; these are the so-called running penumbral waves, first seen by Giovanelli (1972) and Zirin and Stein (1972).

As well as having modes of oscillation of their own, sunspots interact with the global oscillations of their surroundings. Braun, Duvall and Labonte (1987,1988) have discovered that sunspots appear to be sinks for p-modes, absorbing much of the local power in those modes. Altogether, then, the modes of oscillation of sunspots and the manner in which they interact with the oscillations in their surroundings is a complex topic, not yet fully understood (see Evans and Roberts 1990a for a recent discussion).

Besides sunspots, there are other sources of magnetic field in the photospheric layers of the Sun. In fact, there appears to be a hierarchy of magnetic structures, ranging from the large-scale sunspots with diameters of some 10^4 km, to pores and knots on a scale of 10^3 km, through to the smallest elements, the intense flux tubes. Outside of sunspots, intense flux tubes (or faculae) carry over 90 percent of the magnetic flux in the photospheric layers of the Sun (see the

review by Stenflo 1989). They have diameters of about 200 km and field strengths of about 0.2T (= 2 kG). Thus, we have the remarkable fact that virtually all of the magnetic flux coming through the solar surface is found to be in concentrated forms.

What is the nature of an intense flux tube, and what types of waves may it support? The very fact that the magnetic field is confined to a thin tube offers the possibility of a theoretical description, a *thin flux tube theory*, incorporating the effects of both magnetism and stratification (see the reviews by Spruit 1981; Spruit and Roberts 1983; Thomas 1985; Roberts 1986,1990a,b; Ryutova 1990a,b; Schussler 1990). Observationally, the small-scale nature of an intense tube, smaller than the currently best available telescopic resolution, makes any detection of wave motions, expected on theoretical grounds, problematic. Nonetheless, Solanki and Roberts (1990) have argued that certain Stokes asymmetries detected in spectroscopic lines formed within tubes find a natural explanation if waves are present within the tubes.

Moving higher in the solar atmosphere we come to the corona. Here the plasma β is small because the gas pressure is so low in this hot but tenuous, near vacuum, region. Consequently, although magnetic field strengths are estimated to be substantially less than those measured in sunspots and intense flux tubes - Zeeman splitting cannot generally be measured in the tenuous corona - magnetic effects are nonetheless dominant. It is natural to believe, then, that the coronal plasma is capable of supporting a range of oscillatory phenomena. Indeed, it is conjectured that magnetohydrodynamic waves may be responsible for maintaining the $2 \times 10^6 K$ closed-field, active region corona and the $1 \times 10^6 K$ open-field regions known as coronal holes. The heating of the corona, however, remains a puzzle : of whether it is due to the dissipation of magnetohydrodynamic waves (acoustic waves are ruled out), or through a multiple series of mini-reconnections (nano-flares) attempting to relieve stresses within the field (built up by photospheric motions), or yet some other mechanism. For recent discussions of coronal heating, see Hollweg (1990a,b), Davila (1990) and Parker (1987, 1990).

Aside from the question of heating, radio observations of the corona have revealed a number of pulsation phenomena (see, for example, Aschwanden 1987) which may be indicative of waves. The observed inhomogeneous nature of the corona suggests, on theoretical grounds, that the modes of oscillation it can support may be very different from those of a uniform medium. Indeed, Edwin and Roberts (1983) and Roberts, Edwin and Benz (1984) have shown theoretically that a low-β corona may support rapid oscillations which are ducted in regions of low Alfvén speed (see also Meerson, Sasarov and Stepanov 1978; Habbal, Leer and Holzer 1979). These guided waves are mathematically akin to Pekeris waves in ocean layers and Love waves in the Earth's crust, earning them the names of *magnetic Love* and *magnetic Pekeris* waves (Roberts, Edwin and Benz 1984).

Taken overall, this brief outline of oscillatory phenomena observed in the Sun serves perhaps to show that the Sun is indeed a complex medium capable of supporting a large variety of wave phenomena. The question of how we may attempt to describe such phenomena is taken up next.

6.2 Magnetohydrodynamic Waves

6.2.1 Characteristic Speeds

We are familiar with the propagation of sound in an ideal gas (see, for example, Lighthill 1978). Compressions and rarefactions in the atmosphere are propagated isotropically with the sound speed c_s, defined in terms of the undisturbed gas pressure p_0 and density ρ_0 by

$$c_s = (\gamma p_0/\rho_0)^{1/2}, \tag{6.2}$$

where γ is the ratio of specific heat. An initiated disturbance, such as from a clap of the hands, results in a wave which propagates radially outward from the source with speed c_s, the gas undergoing variations in pressure, density and temperature with motions in the gas being aligned with the direction of wave propagation.

In the presence of an applied magnetic field variations in gas pressure will generally lead to disturbances of the magnetic field-lines. In a perfect conductor, the magnetic field-lines and the fluid motions are frozen together, so any attempt to initiate a sound wave will result in variations in the magnetic field, as it is locally compressed or rarefied. As a result, sound is no longer able to propagate with the speed c_s, and the directionality of the applied magnetic field renders wave propagation anisotropic.

The characteristic speed of a magnetic disturbance is the Alfvén speed c_A, defined by

$$c_A = \left(\frac{B_0^2}{\mu\rho_0}\right)^{1/2} \tag{6.3}$$

for a magnetic field of strength B_0 and magnetic permeability μ, embedded in a gas of density ρ_0. We may note that the plasma β, introduced earlier, is related to the sound and Alfvén speeds by

$$\beta = \frac{p_0}{\left(\frac{B_0^2}{2\mu}\right)} = \frac{2}{\gamma}\frac{c_s^2}{c_A^2}. \tag{6.4}$$

Generally, we will take $\gamma = 5/3$ for solar applications and so equation (6.4) reveals that β is roughly the square of the ratio of the sound speed to the Alfvén speed. A low-β plasma, such as the corona, is thus one for which the Alfvén speed greatly exceeds the sound speed, $c_A >> c_s$. Wave propagation, then, involves the two speeds c_s and c_A. In fact, the sound speed does not enter into a description of propagation speeds on its own but only in combination with the Alfvén speed. The Alfvén speed, however, does occur on its own. The combined speed (squared) arising from c_s^2 and c_A^2 are constructed thus:

$$c_f^2 = c_s^2 + c_A^2, \qquad c_T^{-2} = c_s^{-2} + c_A^{-2}. \tag{6.5}$$

The larger of the two speeds introduced in equation (6.5), namely c_f, will be referred to as the *fast* (magnetoacoustic) speed; it is both super-sonic and super-Alfvénic. By contrast, the *slow* speed c_T is both sub-sonic and sub-Alfvénic; it arises in the description of waves in a magnetic flux tube (hence the designation).

We may illustrate these speeds numerically. In the corona, for example, the sound speed is about 200 km s^{-1} and the Alfvén speed is perhaps 1000 km s^{-1}, giving a fast speed of c_f = 1020 km s^{-1} and a slow speed of c_T = 196 km s^{-1}. Lower in the solar atmosphere, in an intense photospheric flux tube, for example, the sound speed and Alfvén speed are roughly comparable; with $c_s = c_A$= 10 km s^{-1} we obtain c_f= 14 km s^{-1} and c_T= 7 km s^{-1}.

6.2.2 Waves in a Structured Medium

Consider a basic state in which there is an applied unidirectional magnetic field $B_0\,\hat{\mathbf{z}}$ aligned with the z-axis in an appropriate coordinate system. The equilibrium gas density ρ_0, the pressure p_0, and the field strength B_0 may all vary in a direction perpendicular to the applied magnetic field. This treatment includes, of course, the more elementary case of a *uniform* medium (for which B_0, p_0 and ρ_0 are simply constants).

The governing system of differential equations are those of ideal magnetohydrodynamics (see Chapter One). We consider isentropic

disturbances. Our equations are: the equation of continuity,

$$\frac{\partial \rho}{\partial t} + div\,\rho\mathbf{v} = 0; \tag{6.6}$$

the momentum equation,

$$\rho\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot grad\right)\mathbf{v} = -grad\,p + \mathbf{j}\times\mathbf{B} + \mathbf{g}\rho; \tag{6.7}$$

the induction equation of an electrically ideal fluid,

$$\frac{\partial \mathbf{B}}{\partial t} = curl(\mathbf{v}\times\mathbf{B}); \tag{6.8}$$

Ampere's relation,

$$\mu\mathbf{j} = curl\ \mathbf{B}; \tag{6.9}$$

the solenoidal constraint,

$$div\ \mathbf{B} = 0; \tag{6.10}$$

the equation for isentropic disturbances,

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot grad\right)p = \frac{\gamma p}{\rho}\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot grad\right)\rho; \tag{6.11}$$

and, finally, the ideal gas law,

$$p = \frac{k_B}{\hat{m}}\rho T. \tag{6.12}$$

Here, p, ρ and T denote the pressure, density and temperature of the ideal gas, k_B denotes Boltzmann's constant, $\hat{m}$ is the mean particle mass, and γ is the ratio of specific heats. The magnetic (induction) field is $\mathbf{B}$ and the current density is $\mathbf{j}$. The fluid motion is $\mathbf{v}$, and $\mathbf{g}$ denotes the local gravitational acceleration (of magnitude 0.27 km s^{-2} in the solar atmosphere).

In this Section the effects of stratification will be ignored. Then, with $g = 0$, the equilibrium ($\mathbf{v} = 0$) state of a structured unidirectional magnetic field is one of constant total (gas plus magnetic) pressure:

$$grad\left(p_0 + \frac{B_0^2}{2\mu}\right) = 0. \tag{6.13}$$

To examine disturbances about this equilibrium we consider the linearized form of equations (6.6)-(6.12). It proves convenient to introduce the perturbation in total pressure, p_T:

$$p_T = p + \frac{1}{\mu}\mathbf{B}_0\cdot\mathbf{B}. \tag{6.14}$$

Then, splitting the velocity field $\mathbf{v}$ into components perpendicular

and parallel to the applied magnetic field, equations (6.6)-(6.12) yield

$$\rho_0\left(\frac{\partial^2}{\partial t^2} - c_A^2\frac{\partial^2}{\partial z^2}\right)\mathbf{v}_\perp + \boldsymbol{\nabla}_\perp\left(\frac{\partial p_T}{\partial t}\right) = 0, \tag{6.15}$$

$$\rho_0\left(\frac{\partial^2}{\partial t^2} - c_T^2\frac{\partial^2}{\partial z^2}\right)v_z + \left(\frac{c_s^2}{c_f^2}\right)\frac{\partial}{\partial z}\left(\frac{\partial p_T}{\partial t}\right) = 0, \tag{6.16}$$

where pressure variations are related to the flow through

$$\frac{\partial p_T}{\partial t} = \rho_0 c_A^2\frac{\partial v_z}{\partial z} - \rho_0 c_f^2 div\ \mathbf{v}. \tag{6.17}$$

Here we have written $\mathbf{v} = \mathbf{v}_\perp + v_z\hat{\mathbf{z}}$, and $\boldsymbol{\nabla}_\perp$ denotes the component of the gradient operator that is perpendicular to the direction of the applied magnetic field (i.e., perpendicular to the z-axis). Observe the presence of all four characteristic speeds, c_s, c_A, c_T and c_f, each a function of the coordinate perpendicular to the applied magnetic field.

The system (6.15)-(6.17) governs the behaviour of the four unknowns, namely, the three components of the flow and the associated magnetic perturbation. Notice how the anisotropy introduced by the directionality of the applied magnetic field is reflected in these equations: the direction perpendicular to the magnetic field - equation (6.15) - involves the Alfvén speed, whereas the direction parallel to the field (equation (6.16)) introduces the tube speed c_T. We may note that equation (6.16), for the flow along the applied field, may be written in the simpler form

$$\frac{\partial^2 v_z}{\partial t^2} = c_s^2\frac{\partial}{\partial z}(div\ \mathbf{v}). \tag{6.18}$$

Notice first the reduction of our system (6.15)-(6.17) in the absence of a magnetic field. With $c_A = 0$, we obtain $c_T = 0, c_f = c_s$, and $p_T = p$; also the anisotropy between equations (6.15) and (6.16) disappears, as it must. Thus, in the absence of gravity, sound waves in a non-magnetic atmosphere that is structured in density satisfy

$$\rho_0\frac{\partial^2\mathbf{v}}{\partial t^2} + grad\left(\frac{\partial p}{\partial t}\right) = 0, \tag{6.19}$$

with

$$\frac{\partial p}{\partial t} = -\rho_0 c_s^2 div\ \mathbf{v}. \tag{6.20}$$

Hence, we obtain (see also Bergmann 1946)

$$\frac{1}{c_s^2}\frac{\partial^2 p}{\partial t^2} = \rho_0 div\left(\frac{1}{\rho_0}grad\ p\right) \tag{6.21}$$

and

$$\frac{\partial^2\mathbf{v}}{\partial t^2} = c_s^2\nabla^2\mathbf{v}. \tag{6.22}$$

In the special case of a uniform atmosphere (c_s and ρ_0 constants) both p and $\mathbf{v}$ satisfy the standard wave equation.

Returning to the magnetohydrodynamic problem we see that the system (6.15) - (6.17) is considerably more complicated than the acoustic case. We may note, however, one simple solution, namely, an incompressible motion ($div\ \mathbf{v} = 0$) involving no flow along the field (so $v_z = 0$) and no total pressure variations (so $p_T = 0$). Then equations (6.15) - (6.17) are satisfied by a flow $\mathbf{v}_\perp$ given by

$$\frac{\partial^2\mathbf{v}_\perp}{\partial t^2} = c_A^2\frac{\partial^2\mathbf{v}_\perp}{\partial z^2}. \tag{6.23}$$

In other words, motions in the plane perpendicular to the applied field are solutions of the one-dimensional wave equation, with general solution

$$\mathbf{v}_\perp = \mathbf{f}_\perp(z - c_A t) + \mathbf{g}_\perp(z + c_A t) \tag{6.24}$$

for arbitrary functions $\mathbf{f}_\perp$ and $\mathbf{g}_\perp$ of coordinates $(z - c_A t)$ and $(z + c_A t)$, respectively; $\mathbf{f}_\perp$ and $\mathbf{g}_\perp$ have arbitrary dependence on the coordinates perpendicular to the field. This is the *Alfvén wave*, the transverse vibration of each field-line.

Finally, we end this Section by presenting the reduction of equations (6.15)-(6.17) in Fourier components. For example, in a cartesian coordinate system (x, y, z), with $B_0 = B_0(x)$ and $\rho_0 = \rho_0(x)$, we may Fourier analyse equations (6.15)-(6.17) by writing

$$v_x(x, y, z, t) = v_x(x)\exp i(\omega t - k_y y - k_z z), \tag{6.25}$$

with similar expressions for other variables. Here ω is the angular frequency of the perturbation, and k_y and k_z are the wavenumbers. Then, equations (6.15)- (6.17) reduce to the single ordinary differential equation (Goedbloed 1971; Chen and Hasegawa 1974; Roberts 1981)

$$\frac{d}{dx}\left\{\frac{\rho_0(x)(k_z^2c_A^2(x) - \omega^2)}{m^2(x) + k_y^2}\frac{dv_x}{dx}\right\} = \rho_0(x)(k_z^2c_A^2(x) - \omega^2)v_x, \tag{6.26}$$

where

$$m^2 = \frac{(k_z^2c_s^2 - \omega^2)(k_z^2c_A^2 - \omega^2)}{(c_s^2 + c_A^2)(k_z^2c_T^2 - \omega^2)} \tag{6.27}$$

and may be positive or negative.

Similarly, for a cylindrical coordinate system (r, θ, z), with $B_0 =$

$B_0(r)$ and $\rho_0 = \rho_0(r)$, we may write

$$v_r(r,\theta,z,t) = v_r(r)\exp i(\omega t + n\theta - k_z z) \tag{6.28}$$

for mode number $n = 0, 1, 2, \ldots$. Then equations (6.15)-(6.17) yield (Hain and Lust 1958; Goedbloed and Hagenbeuk 1972; Edwin and Roberts 1983:Roberts 1986)

$$\frac{d}{dr}\left\{\frac{\rho_0(r)(k_z^2c_A^2(r)-\omega^2)}{(m^2(r)+\frac{n^2}{r^2})}\,\frac{1}{r}\,\frac{d}{dr}(rv_r)\right\} = \rho_0(r)(k_z^2c_A^2-\omega^2)v_r, \tag{6.29}$$

with m^2 defined by equation (6.27).

Solution of equations (6.26) and (6.28) is not straightforward; indeed, only a few special cases have been analysed. For one thing these equations possess continuous spectra, a discussion of which would take us too far afield (see Goedbloed 1983; Adam 1982, 1986). Also, investigations of coronal heating by waves are commonly built upon these equations (see Davila 1990; Hollweg 1990 a,b; Goossens 1991). We content ourselves in what follows with the simpler case of two uniform media, such as arises at a magnetic interface or in relation to magnetic flux tubes. Both these cases are important for solar studies. But first we have need to treat the simple case of a uniform unbounded medium, to which we now turn.

6.2.3 Waves in a Uniform Medium

The special case of a uniform medium (for which ρ_0, c_A, c_s, c_T and c_f are all constants) follows immediately from the more general treatment in Section 6.2.2. We readily obtain from equations (6.15)-(6.17) the result (Lighthill 1960)

$$\frac{\partial^4\Delta}{\partial t^4} - (c_s^2+c_A^2)\frac{\partial^2}{\partial t^2}\nabla^2\Delta + c_s^2c_A^2\frac{\partial^2}{\partial z^2}\nabla^2\Delta = 0, \tag{6.30}$$

where $\Delta \equiv div\ \mathbf{v}$. Fourier analysis then yields the dispersion relation

$$\omega^4 - (c_s^2+c_A^2)k^2\omega^2 + c_s^2c_A^2k_z^2k^2 = 0, \tag{6.31}$$

for wavevector $\mathbf{k} = (k_x, k_y, k_z)$ with magnitude $k = (k_x^2+k_y^2+k_z^2)^{1/2}$.

The dispersion relation (6.31) applies to modes that are compressive ($\Delta \neq 0$); these are the *magnetoacoustic waves*, designated as *slow* or *fast* in accordance with the relative magnitude of the two roots of the quadratic equation (6.31) for ω^2.

If $\Delta = 0$, then incompressible disturbances may arise; these are the *Alfvén waves*, noted earlier, and have dispersion relation

$$\omega^2 = k_z^2c_A^2. \tag{6.32}$$

Considering the modes of a uniform medium from the view-point

of the differential equation (6.26) gives immediately the result that

$$(k_z^2c_A^2-\omega^2)(m^2+k_x^2+k_y^2)v_x = 0. \tag{6.33}$$

Thus, for $v_x \neq 0$, we have the Alfvén wave (with $\omega^2 = k_z^2c_A^2$) and the magnetoacoustic waves with

$$m^2 + k_x^2 + k_y^2 = 0. \tag{6.34}$$

Equation (6.34) is, of course, relation (6.31) in disguise. But the form (6.34) that arises naturally from the differential equation makes clear that $m^2 < 0$ for the magnetoacoustic modes of an unbounded medium (for which $k_x^2, k_y^2 > 0$ applies); thus ω^2/k_z^2 lies either between c_T^2 and the smaller of c_s^2 and c_A^2 (corresponding to the *slow magnetoacoustic wave*) or is greater than the larger of c_s^2 and c_A^2 (corresponding to the *fast magnetoacoustic wave*).

The properties of the three waves of a uniform medium are readily obtained from the discussion we have presented and are in any case documented elsewhere (see, for example, Cowling 1976; Roberts 1985a), so we will pass on to deal with other matters.

6.3 Surface Waves

A medium with a sharp change in physical properties across an interface has the ability to support surface waves, modes that propagate along the interface. To discuss this possibility we return to equations (6.15)-(6.17) restricted, for simplicity, to the case of parallel propagation ($\partial/\partial y = 0$) and two-dimensional motions $\mathbf{v} = (v_x, 0, v_z)$. Then

$$\frac{d}{dx}\left(\frac{\rho_0c_f^2(k_z^2c_T^2-\omega^2)}{(k_z^2c_s^2-\omega^2)}\,\frac{dv_x}{dx}\right) = \rho_0(k_z^2c_A^2-\omega^2)v_x, \tag{6.35}$$

with v_x and p_T related by

$$p_T(x) = \frac{i\rho_0}{\omega}c_f^2\left(\frac{k_z^2c_T^2-\omega^2}{k_z^2c_s^2-\omega^2}\right)\frac{dv_x}{dx}. \tag{6.36}$$

Suppose that the density, pressure, temperature and magnetic field strength change discontinuously across the surface $x = 0$. Set

$$B_0(x) = \begin{cases} B_e, & x > 0, \\ B_0, & x < 0, \end{cases} \qquad \rho_0(x) = \begin{cases} \rho_e, & x > 0, \\ \rho_0, & x < 0, \end{cases} \tag{6.37}$$

for constants B_0, B_e, ρ_0 and ρ_e. Then pressure balance (equation (6.13)) requires that

$$p_0 + \frac{B_0^2}{2\mu} = p_e + \frac{B_e^2}{2\mu}, \tag{6.38}$$

which, when combined with the ideal gas law, gives

$$\frac{\rho_e}{\rho_0} = \frac{c_0^2 + \frac{1}{2}\gamma c_{A0}^2}{c_e^2 + \frac{1}{2}\gamma c_{Ae}^2}. \tag{6.39}$$

Quantities with a suffix '0' refer to the region $x < 0$, those with a suffix 'e' to the region $x > 0$.

In each of the regions $x > 0$ and $x < 0$ the medium is uniform and so equation (6.35) applies with its coefficients constants:

$$\frac{d^2 v_x}{dx^2} + m^2 v_x = 0, \tag{6.40}$$

with m^2 defined by equation (6.27) and equal to m_0^2 in $x < 0$, m_e^2 in $x > 0$.

We require a solution such that $\rho_0(x) v_x^2 \to 0$ as $|x| \to \infty$, corresponding to the energy of the motion being confined to near the interface. Thus

$$v_x(x) = \begin{cases} A_e e^{-m_e x}, & x > 0, \\ A_0 e^{m_0 x}, & x < 0. \end{cases} \tag{6.41}$$

In writing the solution (6.41) we must require that m_0 and m_e be positive.

Across the interface $x = 0$ we require that $v_x(x)$ be continuous and that $p_T(x)$ be continuous also, this latter condition being apparent from the integral of equation (6.35) across the interface. Thus, $A_0 = A_e$ and

$$\rho_0(k_z^2 c_{A0}^2 - \omega^2) m_e + \rho_e(k_z^2 c_{Ae}^2 - \omega^2) m_0 = 0. \tag{6.42}$$

This is the dispersion relation describing the parallel propagation of *magnetoacoustic surface waves* at a single interface (Wentzel 1979; Roberts 1981). Notice that it is transcendental.

One property of surface waves follows immediately from equation (6.42), namely that the phase-speed ω/k_z must lie between the Alfvén speeds, c_{A0} and c_{Ae}, of the two media. In fact there are, in general, *two* surface waves, which we may refer to as *slow* and *fast surface waves* (Roberts 1981). The waves are compressive ($div\ \mathbf{v} \neq 0$) and so are *magnetoacoustic* surface waves. (The common term 'Alfvénic surface wave' is a misnomer.) The conditions for their existence have been explored in Roberts (1981) and Miles and Roberts (1989). An extension to include the case of non-parallel propagation($k_y \neq 0$) has been carried out in Jain and Roberts (1990).

The simplest case of the dispersion relation (6.42) is that applying in an *incompressible* medium. For with $c_s \to \infty$, we obtain $m^2 \to k_z^2$

and so equation (6.42) reduces to

$$\frac{\omega}{k_z} = \left(\frac{\rho_0 c_{A0}^2 + \rho_e c_{Ae}^2}{\rho_0 + \rho_e}\right)^{1/2}. \tag{6.43}$$

The speed defined by this expression arises in magnetic flux tubes too, as we shall shortly see.

Finally, we note that the *nonlinear* behaviour of magnetoacoustic surface waves may also be investigated (see Ruderman 1985, 1988; Hollweg 1987), though a discussion of these aspects would take us too far afield(see, however, Roberts 1990c).

6.4 Waves in Magnetic Flux Tubes

The importance of magnetic flux tubes in the dynamics of the solar atmosphere raises the question of what waves do they support? To begin with we may note that a flux tube supports a *torsional Alfvén wave*, which corresponds to taking $\mathbf{v} = v_\theta \hat{\theta}$ in a cylindrical coordinate form of equations (6.15)-(6.17). Then $v_\theta(r, z, t)$ satisfies the wave equation

$$\frac{\partial^2 v_\theta}{\partial t^2} = c_A^2 \frac{\partial^2 v_\theta}{\partial z^2}. \tag{6.44}$$

Torsional Alfvén waves have been discussed by Hollweg (1990a,b) and Spruit (1982).

Turning to the general differential equation (6.29) we apply it to a uniform region. Actually, it proves more convenient to solve for p_T (rather than v_r), which satisfies the equation (Edwin and Roberts 1983; Roberts 1986)

$$\rho_0(k_z^2 c_A^2 - \omega^2)\frac{1}{r}\frac{d}{dr}\left\{\frac{1}{\rho_0(k_z^2 c_A^2 - \omega^2)} r \frac{dp_T}{dr}\right\} = \left(m^2 + \frac{n^2}{r^2}\right) p_T. \tag{6.45}$$

For a medium in which ρ_0, c_A and m are constants this is simply Bessel's equation,

$$\frac{d^2 p_T}{dr^2} + \frac{1}{r}\frac{dp_T}{dr} - \left(\frac{n^2}{r^2} + m^2\right) p_T = 0. \tag{6.46}$$

We consider the solution of the above equations under two circumstances, namely that of an isolated magnetic flux tube and that of a non-isolated low-β tube. The first situation is representative of the magnetic flux tubes that occur in the photosphere, including both sunspots and intense tubes. The second case is more appropriate

for coronal conditions, representing in particular coronal loops (i.e., high density flux tubes embedded in a strong magnetic field).

6.4.1 Isolated Tubes

Consider, then, an isolated flux tube of radius a, for which

$$B_0(r) = \begin{cases} B_0, & r < a, \\ 0, & r > a, \end{cases} \qquad \rho_0(r) = \begin{cases} \rho_0, & r < a, \\ \rho_e, & r > a. \end{cases} \tag{6.47}$$

The appropriate solution of equation (6.46) for the region inside the magnetic flux tube is

$$p_T = A_0 I_n(m_0 r), \qquad r < a, \tag{6.48}$$

where I_n denotes a modified Bessel function of order n. There is no restriction on the sign of m_0^2 (defined, for the region $r < a$, by equation (6.27)); modes with $m_0^2 > 0$ have an *evanescent* (non-propagating) character, whereas modes with $m_0^2 < 0$ (for which I_n is better replaced by the Bessel function J_n, applied to argument $|m_0|r$) have an *oscillatory* character. We refer to modes with $m_0^2 > 0$ as *surface* waves, and those with $m_0^2 < 0$ as *body* waves. See Figure 6.1.

The environment of an isolated flux tube is field-free and so governed by equation (6.46) with $c_A = 0$ and p_T now simply the perturbed gas pressure. Thus, requiring that p_T declines to zero far from the tube we select the modified Bessel function K_n as our solution of equation (6.46):

$$p_T = A_e K_n(m_e r), \quad r > a, \tag{6.49}$$

where

$$m_e^2 = k_z^2 - \frac{\omega^2}{c_e^2}. \tag{6.50}$$

The requirement of a declining pressure field outside the tube means that we are imposing the constraint $m_e^2 > 0$, and so $\omega^2 < k_z^2 c_e^2$.

The two solutions (6.48) and (6.49) are matched by requiring, much as in Section 6.3, that the radial velocity component v_r and the pressure perturbation p_T be continuous across $r = a$. The result is the dispersion relation (McKenzie 1970;Roberts and Webb 1978; Wilson 1980; Spruit 1982; Edwin and Roberts 1983; Evans and Roberts 1990a)

$$\rho_e \omega^2 m_0 \frac{I_n'(m_0 a)}{I_n(m_0 a)} + \rho_0 (k_z^2 c_A^2 - \omega^2) m_e \frac{K_n'(m_e a)}{K_n(m_e a)} = 0, \tag{6.51}$$

where the dash denotes the derivative of the Bessel function ($I_n'(x) \equiv dI_n(x)/dx$, etc.).

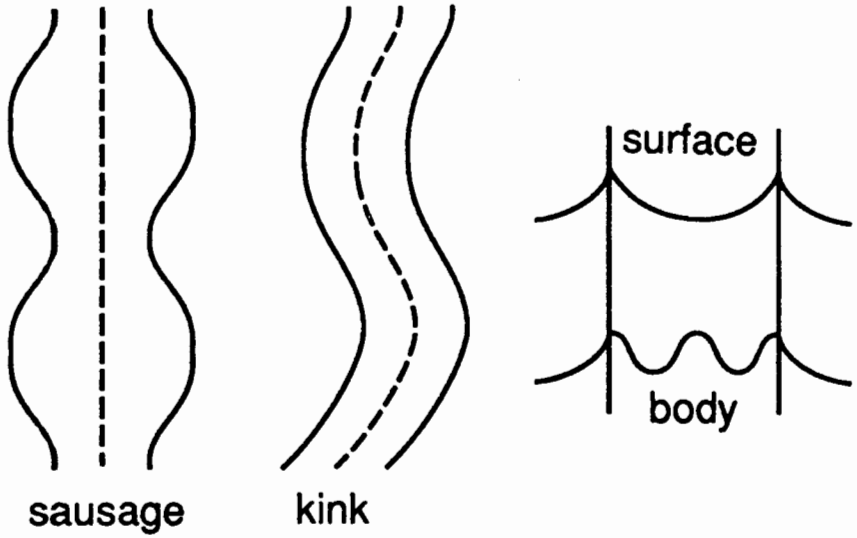


Figure 6.1 The modes of oscillation of a magnetic flux tube.

The two types of solutions of the dispersion relation (6.51), corresponding to the body and surface waves, require a detailed, largely numerical, investigation to extract them from its transcendental structure (see Edwin and Roberts 1983; Evans and Roberts 1990a). Figure 6.2 displays a typical case. The body modes have phase-speeds ω/k_z that are between c_T and c_0 and correspond to slow modes. The *slow body* modes may be viewed as waves that are constrained within the tube, bouncing from side to side of the tube as they propagate along its interior. Body modes have $m_0^2 < 0$ and so are more conveniently obtained from equation (6.51) if we first rewrite it in a form appropriate to that case. With

$$n_0^2 = -m_0^2 = \frac{(\omega^2 - k_z^2 c_A^2)(\omega^2 - k_z^2 c_0^2)}{(c_0^2 + c_A^2)(\omega^2 - k_z^2 c_T^2)}, \tag{6.52}$$

equation (6.51) becomes

$$\rho_e \omega^2 n_0 \frac{J_n'(n_0 a)}{J_n(n_0 a)} + \rho_0 (k_z^2 c_A^2 - \omega^2) m_e \frac{K_n'(m_e a)}{K_n(m_e a)} = 0. \tag{6.53}$$

There are also *slow surface waves*. These are best understood by considering the *incompressible* limit of equation (6.51), obtained by letting c_0 and c_e tend to infinity (formally, $\gamma \to \infty$). The result is (Dungey and Loughhead 1954; Uberoi and Somasundaram 1980)

$$\frac{k_z^2 c_A^2}{\omega^2} = 1 - \left(\frac{\rho_e}{\rho_0}\right) \frac{I_n'(|k_z|a) K_n(|k_z|a)}{I_n(|k_z|a) K_n'(|k_z|a)}. \tag{6.54}$$

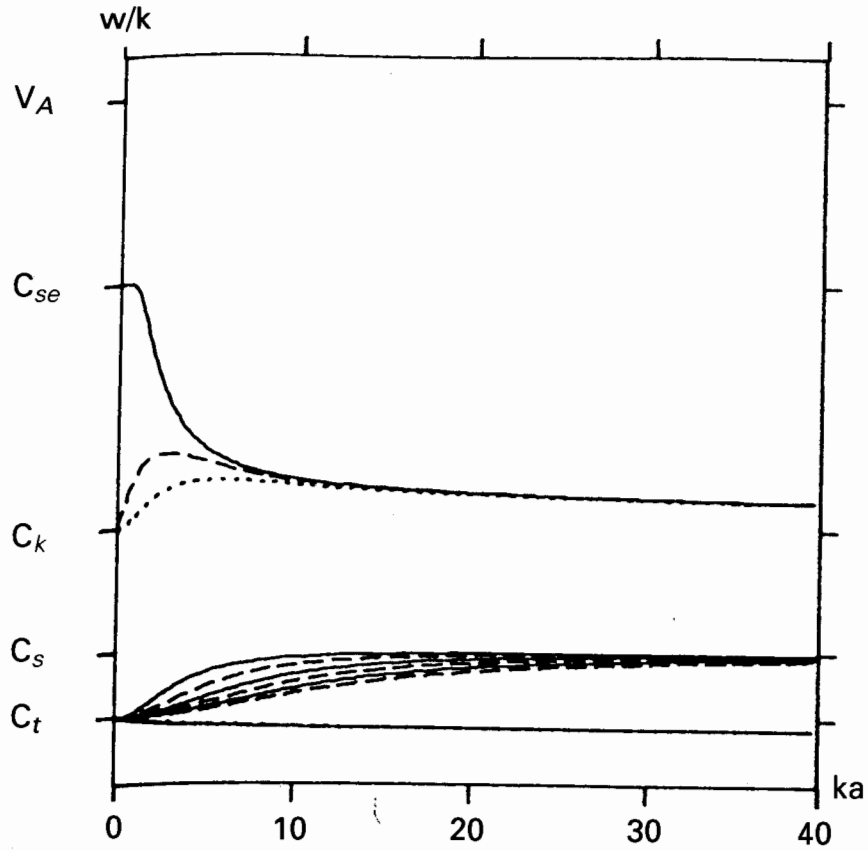


Figure 6.2 The phase-speed as a function of ak_z for waves in a magnetic flux tube. Solid curves are sausage($n = 0$) modes,dashed curves are kink($n = 1$) modes, and dotted curves are $n = 2$ fluting modes. Note: $k \equiv k_z$, $c_{se} \equiv c_e$, $v_A \equiv c_A$ and $c_s \equiv c_0$. (After Edwin and Roberts 1983 and Evans and Roberts 1990a.)

Thus, the transcendental nature of equation (6.51) is removed in the incompressible limit.

It is of interest to examine the *thin tube* limit of equation (6.54). For waves with wavelengths that are much longer than the tube radius, so that $|k_z|a << 1$, we may readily obtain the approximate results:

$$\omega \sim k_z c_A \left[1 - \frac{1}{4}\left(\frac{\rho_e}{\rho_0}\right)k_z^2 a^2 K_0(|k_z|a)\right] \tag{6.55}$$

for the case $n = 0$, and

$$\omega \sim k_z c_k \left[1 + \frac{1}{2}\left(\frac{\rho_e}{\rho_e + \rho_0}\right)k_z^2 a^2 K_0(|k_z|a)\right] \tag{6.56}$$

for the $n = 1$ mode. The speed c_k, defined by

$$c_k^2 = \left(\frac{\rho_0}{\rho_0 + \rho_e}\right)c_A^2, \tag{6.57}$$

is sub-Alfvénic and in fact is precisely the same speed that arises in the description of surface waves on a single magnetic interface (equation (6.43) with $c_{Ae} = 0$).

The occurrence of the Alfvén speed in the above is, of course, entirely expected, given that the medium is incompressible. But it should be noted that the modes we are describing are essentially magnetoacoustic (rather than Alfvénic) in character. The $n = 0$ mode corresponds to symmetric pulsations of the tube and is commonly referred to as the *sausage* mode. The $n = 1$ mode corresponds to serpentine motions of the tube and is usually called the *kink* mode. The $n \geq 2$ waves are *fluting* modes. See Figure 6.1.

Equations equivalent to (6.55) and (6.56) have been obtained for the compressible case too (Edwin and Roberts 1983; Roberts 1985b). The speed c_k for the kink mode survives unscathed but the Alfvén speed c_A in the sausage mode is replaced by the slow magnetoacoustic speed c_T.

The above approximate formulae for the wave speeds have proved important in *nonlinear* theories. Indeed, it has been shown that weakly nonlinear, weakly dispersive waves in the slow sausage surface mode satisfy the integro-differential equation (Roberts 1985b; Molotovshchikov and Ruderman 1987)

$$\frac{\partial v}{\partial t} + c_T\frac{\partial v}{\partial z} + \beta_0 v\frac{\partial v}{\partial z} + \alpha_0\frac{\partial^3}{\partial z^3}\int_{-\infty}^{\infty}\frac{v(s,t)ds}{[\lambda^2 a^2 + (z-s)^2]^{1/2}} = 0, \tag{6.58}$$

where $\lambda = c_T/c_A$ (for $c_0 = c_e$); the constants α_0 and β_0 are complicated coefficients not given here. In the above, $v = v(z,t)$ denotes the flow velocity along the central axis of the tube. Equation (6.58), sometimes referred to as the Leibovich-Roberts equation (Bogdan and Lerche 1988), has been solved numerically (Weisshaar 1990) and exhibits solutions that appear to be soliton-like in character. No analytical solution of equation (6.58) is presently known. The equivalent problem for a *slab* of magnetic field, rather than a cylindrical tube, is known to lead to *solitons*; the governing equation, replacing (6.58), is the Benjamin-Ono equation, namely (Roberts and Mangeney 1982; Roberts 1985b; Merzljakov and Ruderman 1985)

$$\frac{\partial v}{\partial t} + c_T\frac{\partial v}{\partial z} + \beta_0 v\frac{\partial v}{\partial z} + \alpha_1\frac{\partial^2}{\partial z^2}\int_{-\infty}^{\infty}\frac{v(s,t)}{s-z}ds = 0, \tag{6.59}$$

for constant α_1. The Benjamin-Ono equation, involving the Hilbert

transform (with Cauchy principal value integral), has been extensively studied (e.g., Ablowitz and Segur 1981; Matsuno 1984) and indeed has a simple soliton solution.

6.4.2 Coronal Tubes

We turn now to a brief consideration of the case of a magnetic environment. That is, we consider an equilibrium magnetic field and a density of the form

$$B_0(r) = \begin{cases} B_0, & r < a, \\ B_e, & r > a, \end{cases} \qquad \rho_0(r) = \begin{cases} \rho_0, & r < a, \\ \rho_e, & r > a. \end{cases} \tag{6.60}$$

In fact, without loss we could set $B_e = B_0$ provided $\rho_0 \neq \rho_e$ because it is the form of the Alfvén speed that is important. On the face of it this case is similar to that of an isolated tube (given by taking $B_e = 0$), but there are interesting differences in the nature of the modes that arise when the external Alfvén speed, c_{Ae}, is non-zero and in particular exceeds the Alfvén speed c_A within the tube and the two sound speeds, c_0 and c_e.

The derivation of the dispersion relationship proceeds much as in Section 6.4.1, with

$$m_e^2 = \frac{(k_z^2 c_e^2 - \omega^2)(k_z^2 c_{Ae}^2 - \omega^2)}{(c_e^2 + c_{Ae}^2)(k_z^2 c_{Te}^2 - \omega^2)} \tag{6.61}$$

replacing the form given in equation (6.50), to which it reduces when $c_{Ae} = 0$. The resulting dispersion relation is then found to be (Edwin and Roberts 1983)

$$\rho_e(k_z^2 c_{Ae}^2 - \omega^2) n_0 \frac{J_n'(n_0 a)}{J_n(n_0 a)} + \rho_0(k_z^2 c_A^2 - \omega^2) m_e \frac{K_n'(m_e a)}{K_n(m_e a)} = 0, \tag{6.62}$$

with n_0^2 defined in equation (6.52). With $c_{Ae} = 0$ equation (6.62) reduces to equation (6.53), of course.

We are interested in the case of Alfvén speeds exceeding sound speeds, appropriate to the corona. Then equation (6.62) possesses two sets of modes, namely fast and slow body waves (Edwin and Roberts 1983); see Figure 6.3. There are no surface ($n_0^2 < 0$) waves. Both sets of modes are dispersive, with the fast waves being strongly dispersive. The most interesting aspect of these waves is that the fast body waves occur only if $c_{Ae} > c_A$. That is, with $c_A, c_{Ae} > c_0, c_e$, fast body waves occur only for tubes (coronal loops) that satisfy $c_{Ae} > c_A$, and so fast body waves occur in regions of comparatively *low* Alfvén

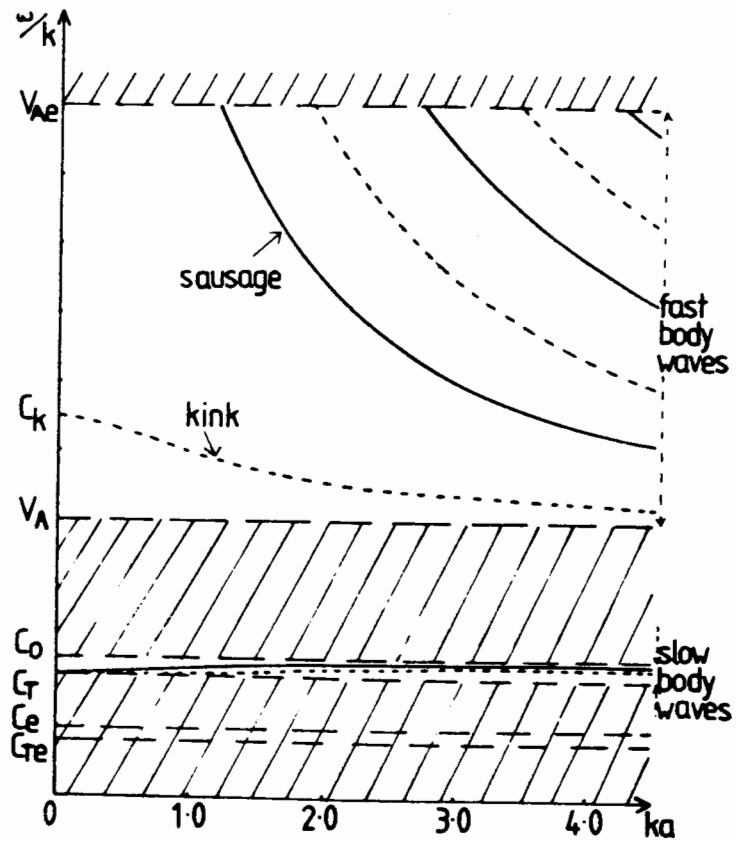


Figure 6.3 The phase-speed ω/k_z of fast and slow body waves in a low-β magnetic cylinder. Note: $k \equiv k_z$, $v_A \equiv c_A$ and $v_{Ae} \equiv c_{Ae}$. (After Edwin and Roberts 1983.)

speed. In the case $B_0 = B_e$ regions of low Alfvén speed correspond to regions of *high* gas density, which is the usual circumstances in coronal loops. In other words, coronal loops are wave guides for fast magnetoacoustic waves (Habbal, Leer and Holzer 1979; Edwin and Roberts 1983). Any attempt to generate fast waves in regions of *high* Alfvén speed would presumably result in those modes being lost from the loop, leaking from its sides (see Spruit 1982; Davila 1985; Cally 1986).

It turns out that there are close analogies here between the behaviour of fast magnetoacoustic body modes, light waves in a fibre optic, and Love waves in the Earth's crust and Pekeris sound waves in an internal ocean layer (Edwin and Roberts 1983; Roberts, Edwin

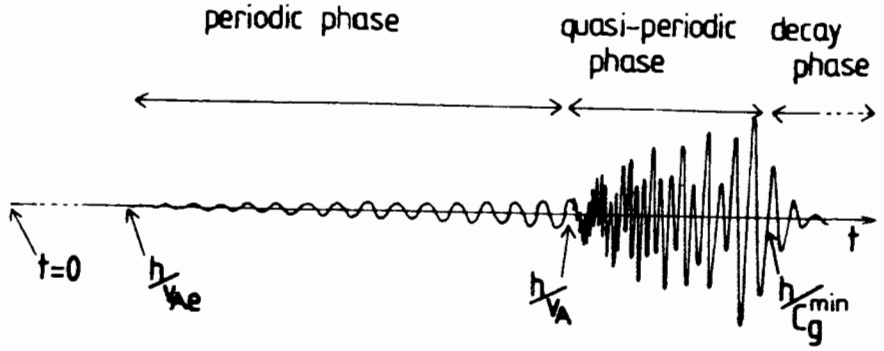


Figure 6.4 The signature of an impulsively generated fast sausage mode, exhibiting three phases of activity. (After Roberts, Edwin and Benz 1984.)

and Benz 1984). Because of the latter analogy, Edwin and Roberts referred to the sausage and kink fast waves as *magnetic Pekeris* and *magnetic Love waves*, respectively.

The analogy between Pekeris and Love waves and fast waves in a coronal loop implies that just as earth-quakes produce distinctive seismic signatures so too should coronal loops. Roberts, Edwin and Benz (1984) argue that such signatures, produced by (say) a solar flare, are in fact evident in the radio pulsations that they produce. The impulsive signature (see Figure 6.4) consists of three parts: a low amplitude *periodic phase*, followed by a larger amplitude *quasi-periodic phase*, and finally a *decay* (Airy) *phase*. The shortest time-scales in the motions are those in the quasi-periodic phase. An estimate of the time-scale in the periodic phase is provided by

$$\tau = \frac{2\pi a}{j_0 c_A}, \tag{6.63}$$

where j_0 (=2.40) denotes the first zero of the Bessel function J_0. For

typical coronal conditions of (say) $a = 10^3$ km and $c_A = 10^3$ km s^{-1} this produces a time-scale of the order of a second or less.

6.5 Stratification Effects

So far in our discussion of magnetohydrodynamic waves we have deferred consideration of the effects of stratification. However, such effects are important, especially in the photosphere and below. We consider such effects in this Section, dividing our attention between localized magnetic structures - the magnetic flux tubes - and global structures are represented by p-modes.

6.5.1 Stratification Effects in Magnetic Flux Tubes

The first effect of gravity on an isolated flux tube is to render it non-uniform in height. The fall-off of the confining gas pressure in the environment of the tube forces the tube to expand outwards. The effect is most easily demonstrated for *thin* tubes, that is for tubes that vary little across their structure.

Considering tubes that are symmetric about the z-axis of a cylindrical coordinate system, we may describe thin tubes by the so-called *thin tube equations*:

$$\frac{\partial}{\partial t}\rho A + \frac{\partial}{\partial z}\rho v A = 0, \tag{6.64}$$

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial z} + g, \tag{6.65}$$

$$\frac{\partial p}{\partial t} + v\frac{\partial p}{\partial z} = \frac{\gamma p}{\rho}\left(\frac{\partial \rho}{\partial t} + v\frac{\partial \rho}{\partial z}\right), \tag{6.66}$$

$$p + \frac{B^2}{2\mu} = p_e, \tag{6.67}$$

$$BA = \text{constant}. \tag{6.68}$$

In equations (6.64)-(6.68), $B(z,t)$ is the field-strength of the thin tube, $A(z,t)$ its cross-sectional area, $v(z,t)$ the longitudinal flow speed within the tube, where the gas pressure and density are $p(z,t)$ and $\rho(z,t)$. The external gas pressure $p_e(z,t)$ is calculated on the boundary of the tube. We have chosen the z-axis to be aligned with gravity, and so point *downwards*. The thin tube equations have been derived in Roberts and Webb (1978) by expanding in Taylor

series the equations of ideal magnetohydrodynamics (equations (6.6)-(6.11)) about the central axis ($r = 0$) of the tube (see also Parker 1974; Defouw 1976; Ferras-Mas, Schussler and Anton 1990).

In the absence of motions ($\mathbf{v} = 0$) the above thin tube equations yield the equilibrium

$$p_0(z) = p_0(0)e^N, \qquad \rho_0(z) = \rho_0(0)\frac{\Lambda_0(0)}{\Lambda_0(z)}e^N, \qquad A_0(z) = A_0(0)e^{N/2}, \qquad B_0(z) = B_0(0)e^{N/2}, \tag{6.69}$$

where

$$N(z) = \int_0^z \frac{dz}{\Lambda_0(z)}, \quad \Lambda_0(z) = \frac{p_0(z)}{g\rho_0(z)} \tag{6.70}$$

and it has been assumed that the pressure scale-height $\Lambda_0(z)$ inside the tube is equal to that outside the tube. The environment is taken to be in hydrostatic equilibrium.

The thin tube equations (6.64)-(6.68) have been studied extensively in the linear regime (see Roberts 1990a,b; Schussler 1990) and there are also some investigations of nonlinear aspects (Ferriz-Mas and Moreno-Insertis 1987; Ferriz-Mas 1988; Ryutova 1990a,b) including numerical simulations (Herbold *et al.* 1985; Musielak *et al.* 1989). Here we simply review briefly the linear aspects.

If we linearise equations (6.64)-(6.68) about the equilibrium (6.69) we obtain (after tedious algebra) the result (Rae and Roberts 1982)

$$\frac{\partial^2 Q}{\partial t^2} - c_T^2(z)\frac{\partial^2 Q}{\partial z^2} + \Omega^2(z)Q = 0. \tag{6.71}$$

We observe here the presence of the tube speed c_T, encountered in Section 6.4.1, and so identify the mode described in equation (6.71) as a slow magnetoacoustic wave. The amplitude $Q(z,t)$ is related to the flow $v(z,t)$ by

$$Q(z,t) = \left[\frac{\rho_0(z)A_0(z)c_T^2(z)}{\rho_0(0)A_0(0)c_T^2(0)}\right]^{1/2} v(z,t), \tag{6.72}$$

and Ω^2 is given by

$$\Omega^2(z) = \frac{c_T^2}{4\Lambda_0^2}\left[3\Lambda_0' + \frac{9}{4} - \frac{2}{\gamma} + 2\beta\left(\frac{\gamma-1}{\gamma} + \Lambda_0'\right)\right], \tag{6.73}$$

the dash denoting differentiation with respect to depth z. In the derivation of equation (6.71) it has been assumed that $p_e(z,t)$ is simply the unperturbed external gas pressure, a function of z but not of time.

Equation (6.71) is of the Klein-Gordon type. It takes its simplest

form for an *isothermal* atmosphere. For with Λ_0 a constant, both c_T and Ω^2 are also constants with Ω^2 reducing to (Defouw 1976; Roberts and Webb 1978)

$$\Omega^2 = \left[\left(\frac{9}{4} - \frac{2}{\gamma}\right) - \left(\frac{3}{2} - \frac{2}{\gamma}\right)\frac{\beta}{\beta + \frac{2}{\gamma}}\right]\omega_a^2, \tag{6.74}$$

where $\omega_a \equiv c_s/2\Lambda_0$ is the acoustic cut-off frequency of an isothermal atmosphere.

The presence of constant coefficients in the Klein-Gordon equation (6.71) applied to an isothermal atmosphere means that we may immediately write down a dispersion relation for the mode:

$$\omega^2 = k_z^2 c_T^2 + \Omega^2. \tag{6.75}$$

From the dispersion relation (6.75) it is immediately clear that Ω is in fact a cut-off frequency, a generalization of the usual cut-off frequency of an isothermal atmosphere (namely, ω_a) to a vertical magnetic flux tube. A scrutiny of equation (6.74) reveals that the first contribution to Ω^2 on its right handside arises from the geometrical shape of the tube: a *rigid* tube, with exponential cross-section in accordance with the equilibrium (6.69) with $\Lambda_0(z)$ constant, would possess precisely the cut-off frequency $(9/4 - 2/\gamma)^{1/2}\omega_a$. The second contribution is a reflection of the elasticity of the tube, in general reducing the cut-off frequency below that of a rigid ($\beta = 0$) tube.

An impulsively generated wave in the tube would, in accordance with equation (6.75), result in a wave-front propagating with the tube speed c_T. Behind the wave-front an oscillating wake is set in motion, rising and falling with frequency Ω (Rae and Roberts 1982). Such a wake phenomena has been exploited in numerical simulations of a nonlinear disturbance in a rigid tube impinging on the solar transition layer; the result is a spicule-like feature (Hollweg 1982; Sterling and Hollweg 1988), the spicule being a thin jet of gas seen emerging from the chromosphere with a speed of 20-30 km s^{-1} and rising to a height of about 10^4km before fading from view.

Finally, we note that if the atmosphere is non-isothermal, so that $\Lambda_0' \neq 0$, equation (6.71) possesses unstable ($\omega^2 < 0$) solutions. The instability may be in the form of a downdraught within the tube, which results in the tube collapsing in to form a yet stronger magnetic field. The effect is quenched when kilogauss field-strengths are reached (Parker 1978; Webb and Roberts 1978; Spruit 1979; Spruit and Zweibel 1979). This process of *convective collapse* is commonly believed to be an essential ingredient towards an explanation of why the Sun chooses preferentially to arrange its photospheric magnetic

flux into concentrated small-scale flux tubes (Spruit 1981; Spruit and Roberts 1983; Roberts 1984; Thomas 1990).

6.5.2 p- and f-Modes

We end our discussion of stratification effects by examining the influence of thermal stratification on the propagation of sound in a layered atmosphere. This topic is central to the new subject of helioseismology with its quest to learn about the solar interior through a detailed understanding of the oscillations observed with high accuracy at the solar surface.

Consider, then, a plane stratified atmosphere in the absence of any magnetism. The effects of spherical geometry will not be considered. The gas is barometrically stratified, so

$$p_0'(z) = g\rho_0(z). \tag{6.76}$$

Combined with the ideal gas law (eqn (6.12)) this integrates to

$$p_0(z) = p_0(0)exp\left(\int_0^z \frac{dz}{\Lambda_0(z)}\right), \tag{6.77}$$

for pressure scale-height $\Lambda_0(z) = p_0(z)/g\rho_0(z) = k_B T_0(z)/\hat{m}g$.

Linear perturbations about this equilibrium then satisfy the system

$$\frac{\partial \rho}{\partial t} + div\, \rho_0 \mathbf{v} = 0, \tag{6.78}$$

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} = -grad\, p + \rho g \hat{\mathbf{z}}, \tag{6.79}$$

$$\frac{\partial p}{\partial t} + p_0' v_z = c_s^2\left(\frac{\partial \rho}{\partial t} + \rho_0' v_z\right). \tag{6.80}$$

A convenient variable to use in analysing equations (6.78)-(6.80) is the divergence of the velocity field (Lamb 1932):

$$\Delta \equiv div\, \mathbf{v}. \tag{6.81}$$

Writing

$$\mathbf{v} = (v_x(z), 0, v_z(z))\exp i(\omega t - k_x x), \tag{6.82}$$

for angular frequency ω and horizontal wavenumber k_x, equations (6.78)- (6.80) yield the pair of first order differential equations

$$\frac{dv_z}{dz} + \left(\frac{gk_x^2}{\omega^2}\right)v_z = \left(1 - \frac{k_x^2 c_s^2}{\omega^2}\right)\Delta, \tag{6.83}$$

$$c_s^2 \frac{d\Delta}{dz} + g\left(\gamma - \frac{k_x^2 c_s^2}{\omega^2}\right)\Delta = \left(\frac{g^2 k_x^2}{\omega^2} - \omega^2\right)v_z. \tag{6.84}$$

Equations (6.83) and (6.84) describe the behaviour of sound waves

in a stratified gas and also buoyancy (gravity) modes. Furthermore, since we have as yet made no assumption about the structure of the sound speed $c_s(z)$, other than γ is constant and the equilibrium satisfies equation (6.76), equations (6.83) and (6.84) also contain the possibility of compressible convection ($\omega^2 < 0$). The sound and gravity modes are commonly denoted as p- and g-modes, a notation going back to Cowling (1941).

In addition to the p- and g-modes of equations (6.83) and (6.84) there is another mode, the f-mode, which is of particular interest. The f-mode has no compression ($\Delta = 0$); its frequency and motion are given by

$$\omega^2 = gk_x, \qquad v_z = e^{-k_x z}. \tag{6.85}$$

Thus, the f-mode is essentially an interface wave, with a frequency that is independent of the stratification (i.e., of the form of $c_s(z)$). The existence (or otherwise) of this mode depends upon the conditions we impose at the top and bottom of the atmosphere. Generally, we shall require that the kinetic energy density, $\frac{1}{2}\rho_0(z)v_z^2$, in the vertical motion tends to zero as $|z| \to \infty$. Such conditions tend to limit the range of wavenumber for which the mode can propagate. We return to this aspect shortly.

Proceeding with equations (6.83) and (6.84), we may eliminate v_z to obtain the second order differential equation

$$\frac{d^2\Delta}{dz^2} + \left(\frac{c_s^{2\prime}}{c_s^2} + \frac{\gamma g}{c_s^2}\right)\frac{d\Delta}{dz} + \left\{\frac{\omega^2 - k_x^2 c_s^2}{c_s^2} - \frac{gk_x^2}{\omega^2}\left(\frac{c_s^{2\prime}}{c_s^2} - (\gamma - 1)\frac{g}{c_s^2}\right)\right\}\Delta = 0. \tag{6.86}$$

This in turn may be cast in the form

$$\frac{d^2 Z}{dz^2} + \kappa^2(z) Z = 0, \tag{6.87}$$

where

$$Z(z) = \left(\frac{\rho_0(z)}{\rho_0(0)}\right)^{1/2}\left(\frac{c_s^2(z)}{c_s^2(0)}\right)\Delta \tag{6.88}$$

and

$$\kappa^2(z) = \frac{\omega^2 - \omega_a^2}{c_s^2} + k_x^2\left(\frac{\omega_g^2}{\omega^2} - 1\right). \tag{6.89}$$

Here ω_a and ω_g are the acoustic cut-off and buoyancy (Brunt-Vaisala) frequencies of a stratified atmosphere, defined by

$$\omega_a^2(z) = \frac{c_s^2}{4H_0^2}(1 + 2H_0'), \qquad \omega_g^2(z) = \frac{g}{H_0} - \frac{g^2}{c_s^2} \tag{6.90}$$

for density scale-height $H_0 \equiv \rho_0/\rho_0'$.

To illustrate the above consider a medium with a linear temperature profile such that

$$c_s^2(z) = c_0^2(1 + \frac{z}{z_0}), \quad z \geq 0, \tag{6.91}$$

for constants c_0 and z_0. Then we may take as our equilibrium the polytrope:

$$\frac{p_0(z)}{p_0(0)} = (1 + \frac{z}{z_0})^{m+1}, \quad \frac{\rho_0(z)}{\rho_0(0)} = (1 + \frac{z}{z_0})^m, \quad z \geq 0, \tag{6.92}$$

where

$$m + 1 = \frac{\gamma g}{(c_s^2)'} \tag{6.93}$$

for polytropic index m. Then, in $z > 0$,

$$H_0(z) = \frac{z + z_0}{m}, \quad \omega_a^2 = \frac{m(m+2)c_0^2}{4z_0(z_0+z)}, \quad \omega_g^2 = [m - \frac{1}{\gamma}(1+m)]\frac{g}{z+z_0}. \tag{6.94}$$

At depth ($z >> z_0$), then, the expression for $\kappa^2(z)$ is dominated by $(\omega^2/c_s^2 - k_x^2)$, and ultimately becomes negative (for $k_x \neq 0$). Thus, the character of the amplitude $Z(z)$ changes from oscillation ($\kappa^2 > 0$) to evanescence ($\kappa^2 < 0$) at some intermediate depth which depends upon ω and k_x. In terms of ray acoustics, a sound wave propagating downwards from the top of the polytrope will suffer refraction and so fail to penetrate deep into the medium.

The form (6.91) of the sound speed permits an exact solution of our equations. In fact, equation (6.86) can be related to Kummer's equation and shown to possess the solution (Lamb 1932)

$$\Delta = e^{-k_x(z+z_0)}\Big[CM(-a_0, m+2, 2k_xz + 2k_xz_0) + DU(-a_0, m+2, 2k_xz + 2k_xz_0)\Big], z > 0, \tag{6.95}$$

for confluent hypergeometric functions M and U (see Abramowitz and Stegun 1967) and arbitrary constants C and D. The parameter a_0 is given by

$$2a_0 = \frac{(m+1)}{\gamma}\frac{\omega^2}{gk_x} + \frac{gk_x}{\omega^2}\Big[(\frac{\gamma-1}{\gamma})m - \frac{1}{\gamma}\Big] - (m+2), \tag{6.96}$$

and is assumed to be non-zero. We require disturbances to be bounded as $z \to +\infty$ and this implies that $C = 0$, since the M function grows exponentially as $z \to +\infty$.

We complete our problem by adding an *isothermal* atmosphere to the region $z < 0$. In an isothermal atmosphere with $c_s = c_0$, the density has an exponential profile:

$$\rho_0(z) = \rho_0(0)exp(\frac{z}{H_0}), \quad H_0 \equiv \frac{c_0^2}{\gamma g}, \quad z < 0, \tag{6.97}$$

where ω_a^2 and ω_g^2 are now constants:

$$\omega_a^2 = \frac{c_0^2}{4H_0^2}, \qquad \omega_g^2 = \frac{g}{H_0}(\frac{\gamma-1}{\gamma}). \tag{6.98}$$

Consequently, equation (6.86) has constant coefficients and possesses the solution

$$\Delta = \Big[\frac{\rho_0(z)}{\rho_0(0)}\Big]^{-1/2} exp(qz), \quad z < 0, \tag{6.99}$$

where, in writing equation (6.99), we have selected the solution that is *least growing with height*, giving a declining kinetic energy density, $\frac{1}{2}\rho_0 v_z^2$, provided that $\kappa^2 < 0$; we have set $q^2 = -\kappa^2$.

Accordingly, the requirement $\kappa^2 < 0$ imposes a cut-off on the allowed range of frequency for given wavenumber k_x. In fact, a sketch of the equation $\kappa^2 = 0$ in the ω, k_x plane produces two zones of propagation, a high frequency zone corresponding to sound waves and a low-frequency one associated with gravity modes. (For a more extensive discussion of this aspect, see, for example, Deubner and Gough (1984) and Roberts (1985a).) The propagation zones are separated by an evanescent zone. When $k_x = 0$, we obtain evanescence ($\kappa^2 < 0$) for $\omega < c_0/2H_0$. For conditions typical of the temperature minimum we may take H_0 =125 km and c_0 =7 km s^{-1}, giving an upper limit in cyclic frequency of $\omega/2\pi$ <4.5 mHz. In fact, p-modes of the Sun are observed to somewhat higher frequencies, possibly as a consequence of chromospheric magnetism.

To link the polytropic and isothermal atmospheres dynamically we require that the vertical component of velocity, v_z, be continuous across $z = 0$. The velocity equation follows immediately from eliminating Δ between equations (6.83) and (6.84):

$$\frac{d}{dz}\Big\{\Big(\frac{\rho_0 c_s^2}{\omega^2 - k_x^2c_s^2}\Big)\frac{dv_z}{dz}\Big\} + \Big\{\rho_0 + \frac{gk_x^2}{\omega^2}\Big(\frac{\rho_0 c_s^2}{\omega^2 - k_x^2c_s^2}\Big)' - \frac{\rho_0 k_x^2 g^2}{(\omega^2 - k_x^2c_s^2)\omega^2}\Big\}v_z = 0, \tag{6.100}$$

an equation that is more complicated in appearance than equation (6.86) that we solved for Δ. With Δ known, of course, we may obtain v_z from equation (6.84). The form of equation (6.100) shows that with v_z continuous we require also, by integration of equation

(6.100) across a thin layer about $z = 0$, that the expression

$$\left(\frac{\rho_0 c_i^2}{\omega^2 - k_x^2 c_s^2}\right)\frac{dv_z}{dz} + \frac{gk_x^2}{\omega^2}\left(\frac{\rho_0 c_s^2}{\omega^2 - k_x^2 c_s^2}\right)v_z$$

be continuous across $z = 0$. (This latter expression is equivalent to $i\omega p + g\rho_0 v_z$.) For the case here, of $c_s(z)$ and therefore $\rho_0(z)$ being continuous, the two matching conditions across $z = 0$ are readily seen to be the equivalent of requiring the continuity of Δ and its derivative $d\Delta/dz$.

Application of these boundary conditions produces the dispersion relation

$$\frac{2U'(-a_0, m+2, 2k_x z_0)}{U(-a_0, m+2, 2k_x z_0)} = 1 + \frac{1}{k_x}\left(q - \frac{1}{2H_0}\right), \qquad (6.101)$$

where $q^2 = -\kappa^2$ and κ^2 is defined by equation (6.89) with ω_a and ω_g given by equation (6.98). The dash denotes the derivative of the U function: $U'(-a_0, m+2, x) \equiv dU(-a, m+2, x)/dx$.

The dispersion relation (6.101) describes the behaviour of modes in a stratified medium possessing a linear temperature profile upon which stands an isothermal atmosphere. It is a complicated relationship, describing, in general, g-modes (becoming convection if $\omega_g^2 < 0$) and p-modes. By looking at the behaviour for small $k_x z_0$, and taking in $z > 0$ an adiabatic atmosphere with $\omega_g^2 = 0$, it may be shown that the p-modes have frequencies close to

$$\omega^2 \approx \left(1 + \frac{2n}{m}\right)gk_x, \qquad (6.102)$$

for $n = 1, 2, 3, \ldots$. These are the parabolic ω, k_x relations observed on the Sun. The corrections to equation (6.102), due to the presence of the isothermal atmosphere, cause a slight reduction in the frequency ω.

The simple model of a polytrope with an isothermal upper atmosphere ('chromosphere') may be extended to include the important effect of a horizontal magnetic field, representing chromospheric magnetism. This has been done by Campbell and Roberts (1989) and Evans and Roberts (1990b), considering various profiles of magnetic field strength. These analyses predict analytically that changes in chromospheric magnetism will result in changes in the frequency-wavenumber relationship for p-modes. It is interesting to note that frequency changes of this type have recently been discovered in the observations (Libbrecht and Woodard 1990).

Another interesting prediction of the chromospheric magnetism calculations is that the f-mode is influenced by magnetic effects and

propagates *faster* than that indicated by the non-magnetic case (for which $\omega^2 = gk_x$). Recent observations by Libbrecht, Woodard and Kaufman (1990) indicate that the f-mode is indeed perturbed from the 'deep water wave' result $\omega^2 = gk_x$, but that it propagates *slower* not faster. This result is presently without explanation.

6.6 Concluding Remarks

The Sun exhibits oscillations of a wide variety, covering both global and local scales. These modes are influenced in varying ways and to various degrees by magnetism. This review has aimed to display some of the features that go into a theoretical description of such wave phenomena, and to indicate the progress towards understanding the Sun's oscillatory behaviour that recent years have brought.

Acknowledgements

I am grateful for the hospitality of the Space Physics Theory Group of the Institute for the Study of Earth, Oceans and Space at the University of New Hampshire, Durham, which afforded me the opportunity to write this article. Particular thanks are due to Professor J V Hollweg. Support of the NASA Space Physics Theory Program under grant NAGW-76 to the University of New Hampshire, Durham, is gratefully acknowledged.

References

Ablowitz, M J & Segur, H (1981) *Solitons and the Inverse Scattering Transform*, SIAM, Philadelphia.

Abramowitz, M & Stegun, I A (1967) *Handbook of Mathematical Functions*, Dover, New York.

Adam, J A (1982) *Physics Reports* **86**, 217.

Adam, J A (1986) *Physics Reports* **142**, 263.

Aschwanden, M J (1987) *Solar Phys.* **111**, 113.

Beckers, J M & Tallant, P E (1969) *Solar Phys.* **7**, 351.

Bergmann, P G (1946) *Acoustical Soc. Amer.* **17**, 329.

Bogdan, T J & Lerche, I (1988) *Quart. Applied Maths.* **XLVI**, 365.

Braun, D C, Duvall, T L & Labonte, B J (1987) *Astrophys. J.* **319**, L27.

Braun, D C, Duvall, T L & Labonte, B J (1988) *Astrophys. J.***335**, 1015.

Cally, P S (1986) *Solar Phys.***103**, 27.

Campbell, W R & Roberts, B (1989) *Astrophys. J.***338**, 538.

Chen, L & Hasegawa, A (1974) *Phys. Fluids* **17**, 1399.

Christensen-Dalsgaard, J, Gough, D O & Toomre, J (1985) *Science* **229**, 923.
Cowling, T G (1941) *Mon. Not. Roy. Astron. Soc.* **101**, 367.
Cowling, T G (1976) *Magnetohydrodynamics*, Adam Hilger, Bristol.
Davila, J M (1985) *Astrophys. J.***291**, 328.
Davila, J M (1990) , in *Basic Plasma Processes on the Sun* (eds. E R Priest & V Krishan), Kluwer:Dordrecht, p.149.
Defouw, R (1976) *Astrophys. J.***209**, 266.
Deubner, F-L (1975) *Astron. Astrophys.* **44**, 371.
Deubner, F-L & Gough, D O (1984) *Ann. Rev. Astron. Astrophys.* **22**, 593.
Dungey, J W & Loughhead, R E (1954) *Aust. J. Phys.* **7**, 5.
Edwin, P M & Roberts, B (1983) *Solar Phys.***88**, 179.
Evans, D J & Roberts, B (1990a) *Astrophys. J.* **348**, 346.
Evans, D J & Roberts, B (1990b) *Astrophys. J.* **356**, 704.
Ferriz-Mas, A (1988) *Phys. Fluids* **31**, 2583.
Ferriz-Mas, A & Moreno-Insertis, F (1987) *Astron. Astrophys.* **179**, 268.
Ferriz-Mas, A, Schussler, M & Anton, V (1990) *Astron. Astrophys.* **210**, 425.
Giovanelli, R G (1972) *Solar Phys.* **27**, 71.
Goedbloed, J P (1971) *Physica* **53**, 412.
Goedbloed, J P (1983) *Lecture Notes in Magnetohydrodynamics*, Rijnhuizen Report 83-145, Assoc. Euratom - FOM, pp.289.
Goedbloed, J P (1984) *Physica* **12D**, 107.
Goedbloed, J P & Hagenbeuk, H J L (1972) *Phys. Fluids* **15**, 1090.
Goossens, M (1991), chapter 7 in this volume.
Habbal, S R, Leer, E & Holzer, T E (1979) *Solar Phys.* **64**, 287.
Hain, K & Lust, R (1958) *Z. Naturforsch.* **13a**, 936.
Herbold, G, Ulmschneider, P, Spruit, H C & Rosner, R (1985) *Astron. Astrophys.* **145**, 157.
Hollweg, J V (1982) *Astrophys. J.* **257**, 345.
Hollweg, J V (1987) *Astrophys. J.* **317**, 918.
Hollweg, J V (1990a) *Computer Phys. Rep.* **12**, 205.
Hollweg, J V (1990b), in *Physics of Magnetic Flux Ropes* (eds C T Russell, E R Priest & L C Lee), AGU : Washington, Geophys. Mono. **58**, p.23.
Jain, R & Roberts, B (1990), in *Mechanisms of Chromospheric and Coronal Heating* (IAU Symp., Heidelberg), edits. P Ulmschneider & E R Priest, Springer Verlag, in press.
Lamb, H (1932) *Hydrodynamics*, Cambridge University Press, Cambridge.
Leibacher, J W & Stein, R F (1971) *Astrophys. Lett.* **7**, 191.
Leibacher, J W, Noyes, R W, Toomre, J & Ulrich, R K (1985) *Scientific Amer.* **253**, 48.
Leighton, R B, Noyes, R W & Simon, G W (1962) *Astrophys. J.* **135**, 474.
Libbrecht, K G & Woodard, M F (1990) *Nature* **345**, 779.
Libbrecht, K G, Woodward, M F & Kaufman, J M (1990) *Astrophys. J. Suppl.*, in press.
Lighthill, J (1960) *Phil. Trans. Roy. Soc.* **A252**, 397.
Lighthill, J (1978) *Waves in Fluids*, Cambridge University Press, Cambridge.
Lites, B W (1988) *Astrophys. J.* **334**, 1054.
Matsuno, Y (1984) *Bilinear Transformation Method*, Academic Press, New York.

McKenzie, J F (1970) *J. Geophys. Res., Space Phys.* **75**, 5331.
Meerson, B I, Sasarov, P V & Stepanov, A V (1978) *Solar Phys.* **58**, 165.
Merzljakov, E G & Ruderman, M S (1985) *Solar Phys.* **95**, 51.
Miles, A J & Roberts, B (1989) *Solar Phys.* **119**, 257.
Molotovshchikov, A L & Ruderman, M S (1987) *Solar Phys.* **109**, 247.
Moore, R & Rabin, D (1985) *Ann. Rev. Astron. Astrophys.* **23**, 239.
Musielak, Z E, Rosner, R & Ulmschneider, P (1989) *Astrophys. J.* **337**, 470.
Parker, E N (1974) *Astrophys. J.* **189**, 563.
Parker, E N (1978) *Astrophys. J.* **221**, 368.
Parker, E N (1987) *Physics Today* **40**, No.7, 36.
Parker, E N (1990), in *Physics of Magnetic Flux Ropes* (eds. C T Russell, E R Priest & L C Lee), AGU: Washington, Geophys. Mono. **58**, p.195.
Rae, I W & Roberts, B (1982) *Astrophys. J.* **256**, 761.
Roberts, B (1981) *Solar Phys.* **69**, 27.
Roberts, B (1984) *Advances Space Research* **4**,17.
Roberts, B (1985a), in *Solar System Magnetic Fields* (ed. E R Priest), Reidel: Dordrecht, Chap. 3.
Roberts, B (1985b) *Physics Fluids* **28**, 3280.
Roberts, B (1986), in *Small-Scale Magnetic Flux Concentrations in the Solar Photosphere* (eds. W Deinzer, M Knolker & H Voight), Vandenhoeck & Ruprecht, Gottingen, p.169.
Roberts, B (1990a), in *Physics of Magnetic Flux Ropes* (eds. C T Russell, E R Priest & L C Lee), AGU: Washington, Geophys. Mono. **58**, p.113.
Roberts, B (1990b), in *Basic Plasma Processes on the Sun* (eds. E R Priest & V Krishan), Kluwer:Dordrecht, p.159.
Roberts, B (1990c), in *Mechanics of Chromospheric and Coronal Heating* (eds. P Ulmschneider & E R Priest), in press.
Roberts, B, Edwin, P M & Benz, A O (1984) *Astrophys. J.* **279**, 857.
Roberts, B & Mangeney, A (1982) *Mon. Not. Roy. Astron. Soc.* **198**, 7p.
Roberts, B & Webb, A R (1978) *Solar Phys.* **56**, 5.
Roberts, B & Webb, A R (1979) *Solar Phys.* **64**, 77.
Ruderman, M S (1985) *Fluid Dynamics* **20**, 85.
Ruderman, M S (1988) *Plasma Phys. and Controlled Fusion* **30**, 1117.
Ryutova, M P (1990a), in *Solar Photosphere: Structure, Convection and Magnetic Fields* (ed. J O Stenflo), IAU Symp. **138**, Reidel, p.229.
Ryutova, M P (1990b), in *Basic Plasma Processes on the Sun* (eds. E R Priest & V Krishan), Kluwer:Dordrecht, p.175.
Schussler, M (1990), in *Solar Photosphere: Structure, Convection and Magnetic Fields* (ed. J O Stenflo) IAU: Reidel, Symp. **138**, p.161.
Solanki, S K & Roberts, B (1990), in *Solar Photosphere: Structure, Convection and Magnetic Fields* (ed. J O Stenflo) IAU: Reidel, Symp. **138**, p.259.
Spruit, H C (1979) *Solar Phys.* **61**, 363.
Spruit, H C (1981) , in *The Sun as a Star* (ed. S Jordan) NASA: Washington, SP-450, p.385.
Spruit, H C (1982) *Solar Phys.* **75**, 3.
Spruit, H C & Roberts, B (1983) *Nature* **304**, 401.
Spruit, H C & Zweibel, E G (1979) *Solar Phys.* **62**, 15.

Stenflo, J O (1989) *Astron. Astrophys. Rev.* **1**, 3.
Sterling, A C & Hollweg, J V (1988) *Astrophys. J.* **327**, 950.
Thomas, J H (1985), in *Theoretical Problems in High Resolution Solar Physics* (ed. H U Schmidt), Max Planck Institute: Munich, MPA 212, p.126.
Thomas, J H (1990), in *Physics of Magnetic Flux Ropes*(eds. C T Russell, E R Priest & L C Lee), AGU: Washington, Geophys. Mono. **58** , p.133.
Uberoi, C (1982) *Solar Phys.* **78**, 351.
Uberoi, C & Somasundaram, K (1980) *Plasma Phys.* **22**, 747.
Ulrich, R K (1970) *Astrophys. J.* **162**, 993.
Webb, A R & Roberts, B (1978) *Solar Phys.* **59**, 249.
Weisshaar, E (1989) *Phys. Fluids* **A1**, 1406.
Wentzel, D G (1979) *Astron. Astrophys.* **76**, 20.
Wilson, P R (1980) *Astron. Astrophys.* **87**, 121.
Zirin, H & Stein, A (1972) *Astrophys. J. Lett.* **178**, L85.

7

Magnetohydrodynamic Waves and Wave Heating in Non-Uniform Plasmas

M. GOOSSENS

Astronomisch Instituut, K.U. Leuven, Belgium

7.1. Introduction

Magnetohydrodynamic (MHD) waves in non-uniform plasmas are of obvious importance for the solar atmosphere. The solar atmosphere is a gigantic plasma which, from the photosphere to the corona, is *strongly structured* by the magnetic field and exhibits *inhomogeneities* in density, pressure and magnetic field on small scales. Each region in the solar atmosphere has its own magnetic plasma structures and observations are giving increasing evidence that many of these magnetic structures support MHD waves. A summary of the different magnetic structures in the solar atmosphere and of the observational information concerning the waves they support can be found in e.g. Edwin and Roberts (1987) and Roberts (1988). In addition to the intrinsic interest in MHD waves for the solar atmosphere, MHD waves may also be important for explaining the heating of the solar corona. The present review will not try to give a complete description of the present status of MHD waves in the solar atmosphere. Instead it will attempt to outline basic properties of MHD waves in nonuniform plasmas and will focus on those aspects of MHD waves that are important for understanding wave heating theories of the solar corona. A wave theory for coronal heating must make sure that the energy requirement of the corona can be met by the observed wave amplitudes and must show that the waves are able to dissipate a considerable part of their energy at the coronal level. As far as the energy requirement is concerned, it seems that observed nonthermal motions can provide sufficient energy to heat the corona (Hollweg, 1983). The great difficulty has been considered to be the low efficiency of viscous and ohmic dissipation, which is very weak in a *uniform plasma* and therefore unable to deposit a substantial

MHD simulations, Proc. Chapman Conf. on the Physics of Magnetic Flux Ropes, Bermuda,(ed C.T. Russell, E.R. Priest and L.C.Lee), p669, 1990.

Otto, A., M. Hesse, and K. Schindler, General magnetic reconnection in 3D systems, to appear in: Proc. IUTAM symposium, Cambridge, UK., 1989.

Parker, E. N. *Cosmical Magnetic Fields*, Oxford University Press, Oxford, 1979.

Priest, E. R., *Solar Flare Magnetohydrodynamics*, Gordon and Breach, New York, 1981.

Priest, E. R., Magnetic reconnection on the sun, in *Magnetic Reconnection in Space andAstrophysical Plasmas*, edited by E. W. Hones, AGU, Washington, 1984.

Priest, E. R., Magnetic energy conversion on the sun, in "Reconnection in Space Plasma," edited by T. D. Guyenne and J. J. Hunt, *ESA Publications,SP-285*, Noordwijk, The Netherlands, p. 73, 1988.

Priest, E. R., and T. G. Forbes, Steady magnetic reconnection in three dimensions, *Solar Phys., 119*, 211, 1989.

Ramati, R., S. A. Colgate, G. A. Dulk, P. Hoyng, J. W. Knight, R. P. Lin, D. B. Melrose, F. Orral, C. Paizis, P. R. Shapiro, D. F. Smith, M. Van Hollebeke, Energetic particles in solar flares, in *Solar flares*, edited by P. A. Sturrock, Colorado Ass. Univ. Press, Boulder, Colorado, 1980.

Schindler, K., A theory of the substorm mechanism, *J. Geophys. Res., 79*, 2803, 1974.

Schindler, K., M. Hesse and J. Birn, General magnetic reconnection, parallel electric fields and helicity, *J. Geophys. Res., 93*, 5547, 1988a.

Schindler, K., M. Hesse, and J. Birn, Aspects of three-dimensional magnetic reconnection, in "Reconnection in Space Plasma," edited by T. D. Guyenne and J. J. Hunt, *ESA Publications,SP-285*, Noordwijk, The Netherlands, p. 5, 1988b.

Song, Y., and R. L. Lysak, Evaluation of twist helicity of flux transfer event flux tubes, *J. Geophys. Res., 94*, 5273, 1989.

Stern, D.P., Euler potentials, *Am. J. Phys., 38*, 494, 1970.

Vasyliunas, V. M., Theoretical models of magnetic field line merging, 1, *Rev. Geophys., 13*, 303, 1975.

Vasyliunas, V. M., An overview of magnetospheric dynamics, in *Magnetospheric Particles and Fields*, edited by B. M. McCormac, p. 99, D. Reidel, Hingham, Mass., 1976.

11

Magnetic Heating of the Solar Corona

M A BERGER

University College London, WC1E 6BT, UK.

11.1 Introduction

The solar atmosphere is extremely hot; typical temperatures are 10^6K, compared with 5×10^3K at the photosphere. Thermal energy must be continually supplied to maintain this temperature against radiative cooling (timescale $\sim$ 1 day). Recent theories of coronal heating invoke magnetic energy dissipation as the source of thermal energy. Two major questions must be answered: how is magnetic energy supplied to the corona, and how is it dissipated?

Coronal physicists distinguish between the open corona, where the magnetic lines of force emerging from the photosphere connect to the interplanetary field, and the closed corona, where field lines form an arch with two endpoints in the photosphere. This review will concentrate on the closed corona. As a further dichotomy, most heating theories can be classified as either wave theories, where Alfvén waves carry energy into the corona; or current dissipation theories, where energy is released from the background magnetic field. Some recent models are hybrids in that they involve MHD waves travelling through a stressed or chaotic background field (Lou and Rosner 1986; Similon and Sudan 1989). This review will concentrate on the current dissipation theories.

Let us start with some general remarks about the coronal magnetic field (see, for example, Priest 1987). First, the corona is highly conducting - consider the dissipation timescale for a structure of size D, $\tau_d(D) = D^2/\eta$ where η is the resistive diffusion coefficient. With $\eta \approx 10^{-6}\,\mathrm{km^2\,sec^{-1}}$, a field variation across a granular length scale of $d_g \sim 10^3$ km (more on this below) will take well over 10,000 years to diffuse away. High conductivity implies that the field lines are frozen into the fluid – each field line is convected with the motions

of the coronal plasma. Furthermore, field lines do not pass through each other; coronal motions preserve the topology of the field (i.e. the pattern in which field lines wind and interlace about each other). These remarks do not apply within reconnection regions, which can develop near sharp gradients in the magnetic field; resistive effects are important in these regions.

Secondly, the corona is a low-beta ($\beta = 8\pi nkT/B^2 \approx 3 \times 10^{-4}$) plasma. This means that magnetic forces dominate over fluid pressures. The opposite holds below the photosphere, where beta is high and kinetic forces dominate. In a low-beta plasma, the equilibrium fields are approximately force-free (i.e. $\vec{J} \times \vec{B} = 0$).

Third, the coronal field is line-tied at the photosphere. The endpoints of coronal field lines respond little to the dynamics of the corona itself. Instead, they are fixed at the photosphere and move with the photospheric flow. The importance of line-tying cannot be overstated. Of all the objects studied in plasma physics and astrophysics, the closed corona provides the clearest example of a completely line-tied MHD system. Line-tying modifies the stability properties of the coronal field (e.g. Velli and Hood 1989); it also prevents magnetic energy from escaping the corona. Trapped magnetic energy forms the basis of the heating theories reviewed here.

Fourth, the Alfvén travel time along a coronal arch ($\sim$ 20 sec) is very short compared to time scales for boundary motions at the photosphere ($\sim 10^3 - 10^5$ sec). When the coronal field is out of equilibrium, the $\vec{J} \times \vec{B}$ Lorentz force generates fluid motions. These motions quickly spread along the entire arch, only to be damped by viscosity and absorption at the photospheric ends of the arch. This process depletes the magnetic energy until the field is close to equilibrium (some amount of fluctuations about the equilibrium will always be present).

The final comment concerns the distribution of magnetic flux at the photosphere. Observations suggest that over 9/10 of the flux is localized into discrete elements (Frazier and Stenflo, 1972). These elements have field strengths of 1-2 kG and radii 200-300 km. Above the photosphere, the volume occupied by the field lines expands as the plasma pressure decreases. Thus in the corona, essentially the entire volume is magnetised (to about 100 G in an active region). The field lines emanating from a single flux element will form a coronal flux tube of typical diameter D roughly equivalent to the typical distance d between photospheric flux elements. In an active region this distance is similar to the granular length scale $d \sim d_g \sim 10^3$ km. Note that such a flux tube may split into several smaller tubes before plunging back into the photosphere. Nevertheless, d provides a natural distance scale for variations in the magnetic field direction.

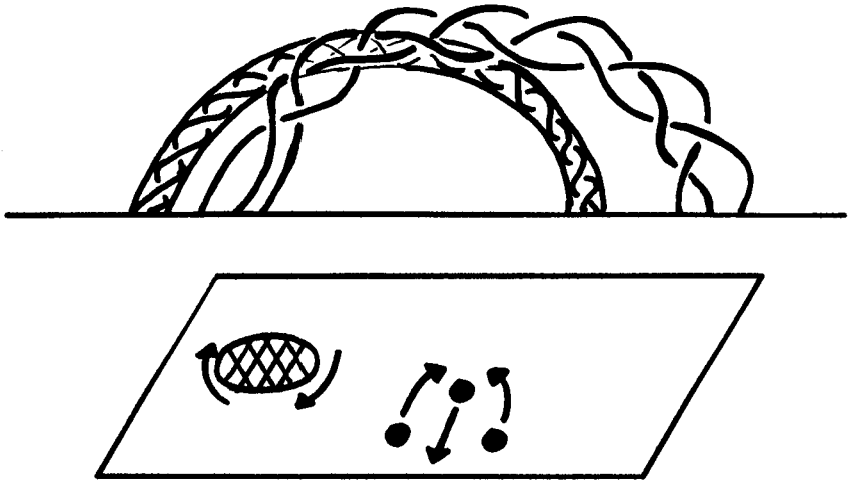
§11.2 discusses in general terms how photospheric motions regenerate the coronal magnetic energy. §11.3 and §11.4 continue this discussion, but with an emphasis on *random* photospheric motions. The Sturrock-Uchida model, involving the random twisting of a single flux tube, is presented in §11.3. §11.4 then covers the Parker model for the random braiding of many flux tubes. §11.5 reviews the Heyvaerts-Priest model for coronal heating, which assumes Taylor relaxation. In §11.6 we look more closely at the dissipation process. In particular, we emphasize the importance of reconnection at thin current layers.

11.2 Generation of Coronal Magnetic Fields

Let us start with a given amount of magnetic flux passing through the photosphere. This flux has been generated deep in the convection zone by processes which need not concern us at the moment. It suffices to assume that $\vec{B} \cdot \hat{n} = 0$ at the photosphere. A simple variational calculation shows that the minimum energy coronal magnetic field, given $\vec{B} \cdot \hat{n}$, is a vacuum (or potential) field where $\nabla \times \vec{B} = 0$. Starting with such a field, we can increase the magnetic energy by moving the endpoints of the magnetic lines at the photosphere. To some extent, waves and fluctuations superimposed on the field increase the energy. However, even if fluid viscosity damped out all the fluctuations, there would still be an increase in the energy of the field.

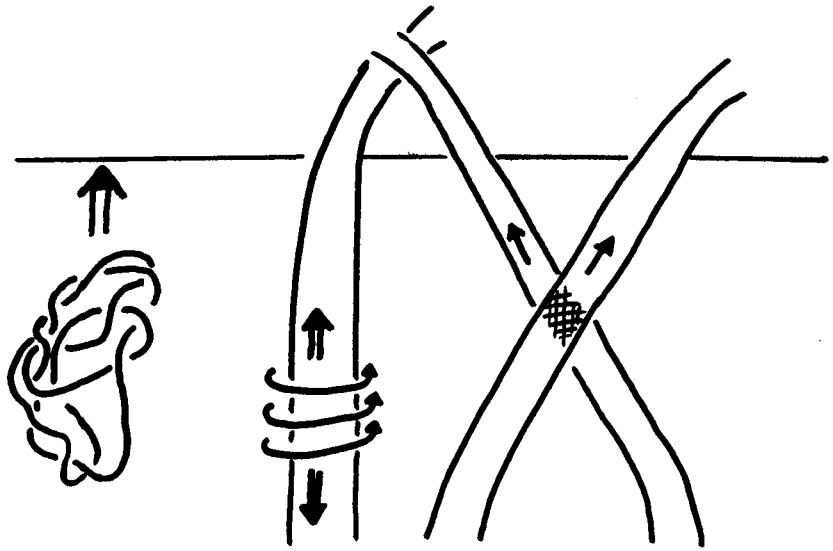
This phenomenon is illustrated in figure 11.1. The twist of the flux rope and the complicated braiding pattern are topological invariants. They cannot be removed from the coronal field except through reconnection (another removal process is mass ejection, where the coronal field and plasma become unconfined and are ejected into the solar wind). Thus if the field settles to a quiescent equilibrium state, that state cannot be the vacuum field, because the vacuum field would have the wrong field topology (i.e. the wrong pattern of twists and braids). The minimum possible energy given a complicated field topology is much greater than that of the vacuum field.

Some processes *below* the photosphere can increase topological



11.1 The lower diagram shows motions at the photosphere. To the left, a vortical motion rotates the footpoint of a coronal arch. This twists the field lines above into a rope; a twisted coronal arch is shown in the background of the upper diagram. Other motions at the photosphere interlace three field line endpoints, resulting in a complicated braiding pattern for the coronal field (foreground of upper drawing).

structure in the corona. Figure 11.2 shows three such processes: on the left, a bubble of magnetized plasma rises through the convection zone. Turbulent motions have endowed the field in the bubble with a complicated structure, which will be carried into the corona as the bubble emerges through the photosphere. The middle drawing shows a single flux tube connecting to the coronal field. Rotational motions in one part of the tube send torsional waves (of opposite signs) up and down the tube. The upward wave will carry twist into the coronal part of the flux tube. Twist in general can travel along a tube via torsional waves; it tends to collect in regions where the plasma pressure is lowest (Parker 1979). The corona has the lowest pressure, so any net twist on a flux tube will try to make its way there. The one trouble with this cartoon is the wave travelling



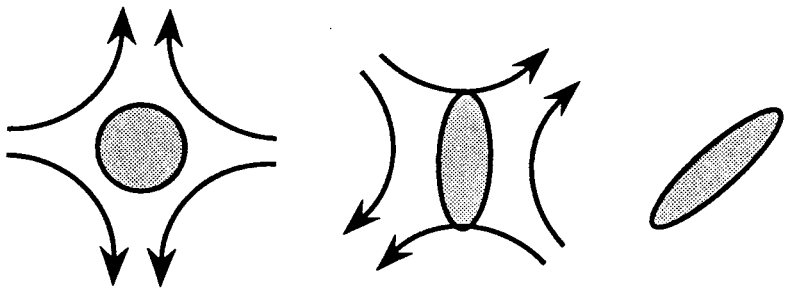
11.2 Magnetic fields below the photosphere.

downwards; it may reflect and travel back upwards. If this wave also reaches the corona, the twist it carries will cancel the twist delivered by the first wave. However, the time delay in traversing a significant part of the convection zone could be days or weeks.

In the right hand drawing, two flux tubes reconnect. By helicity conservation arguments (Song and Lysak 1989; Wright and Berger 1989) reconnection generates precisely one unit of twist (one unit = one complete turn of field lines about the axis of the tube). This is net twist, so there are no problems with cancellation. In general the twist will be equally shared by the two tubes, so the two coronal tubes will each receive 1/2 unit of twist (equivalent to a turn through π).

11.3 The Sturrock-Uchida model

Sturrock and Uchida (1981) considered a single flux tube forming an arch in the solar corona. They argued that the magnetic energy associated with random twisting of the flux tubes would provide a



11.3 Photospheric motions with zero vorticity. A footpoint is first sheared, then rotated.

source for coronal heating. Call the two ends (or footpoints) of the arch at the photosphere D_+ and D_-. These two footpoints move with the surface velocity field. Some of the motion is rotational; after a time t the footpoints will have rotated through angles $\theta_+(t)$ and $\theta_-(t)$. The field lines in the arch will then wind around the axis through an angle $\Theta(t) = -(\theta_+(t) + \theta_-(t))$. For random motions the average winding angle vanishes, $\overline{\Theta}(t) = 0$. However, fluctuations provide a non-zero mean square winding. Assuming that the flows at D_+ and D_- are uncorrelated and have similar statistical properties, the mean square winding angle will be $\overline{\Theta^2}(t) = \overline{\theta_-^2}(t) + \overline{\theta_+^2}(t) = 2\overline{\theta_+^2}(t)$.

We can calculate $\overline{\Theta^2}(t)$ assuming a two-dimensional homogeneous isotropic flow $\vec{V}$ at the photosphere. There are two ways of rotating D_+ (or D_-). First, the vorticity $\omega = \partial_x V_y - \partial_y V_x$ of the flow provides a rotation

$$d\theta_+(t)/dt = \omega/2. \tag{11.1}$$

Letting τ_c be the correlation time of the flow, one finds from vorticity alone

$$d\overline{\theta_+^2}(t)/dt = \tau_c\overline{\omega^2}/2. \tag{11.2}$$

Secondly, even in the absence of vorticity a symmetric shear flow can rotate D_+, provided D_+ has first been sheared into an elongated shape (see figure 11.3). This second process yields a rotation rate approximately half that of the first; the total rotation is thus

$$d\overline{\Theta^2}(t)/dt = 2d\overline{\theta_+^2}(t)/dt \approx 3\tau_c\overline{\omega^2}/2. \tag{11.3}$$

Let $\overline{V^2}$ and λ be the mean square velocity and the correlation length

of the flow. Then

$$\overline{\omega^2} = 4\overline{V^2}/\lambda^2 \tag{11.4}$$

and

$$d\overline{\Theta^2}(t)/dt = 6\tau_c\overline{V^2}/\lambda^2. \tag{11.5}$$

We will consider motions on the granular scale, and set $\lambda = d_g = 10^3\,\mathrm{km}$, $\left(\overline{V^2}\right)^{1/2} = 1\,\mathrm{km\,sec^{-1}}$, and $\tau_c = \lambda/(\overline{V^2})^{1/2} = 800\,\mathrm{sec}$. These values correspond to a diffusion coefficient of $K = 200\,\mathrm{km^2\,sec^{-1}}$, consistent with Smithson's (1973) observations of the jiggling of vertex points of the magnetic network. These observations provide an indication of the random motion of magnetic footpoints due to solar granulation and supergranulation. With the above values $\overline{\Theta^2}(t)$ reaches $(2\pi)^2$ at $t = 1.5$ hours, implying an rms twist of about four turns after 24 hours. Note that additional twisting motions could arise deep in the convection zone; using the photospheric velocity field alone in the calculation may underestimate $\overline{\Theta^2}(t)$. (But we may also be overestimating $\overline{\Theta^2}(t)$ – convective motions are suppressed within intense photospheric flux elements.)

To estimate the energy associated with this twist, consider a cylindrically symmetric flux tube of length L, axial field strength B_z, and radius R. Such a tube has an excess magnetic energy (over that of an untwisted tube) of

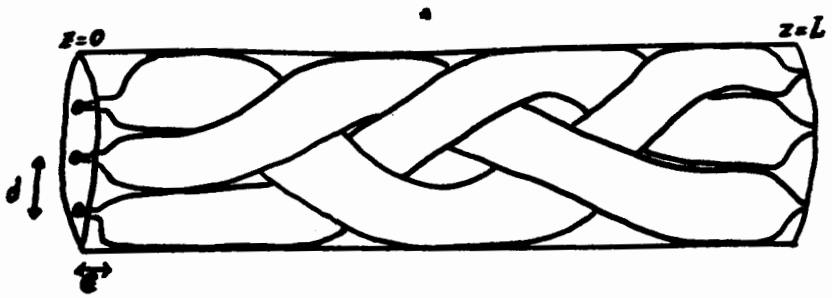
$$\Delta E = \frac{R^4 B_z^2}{16L}\overline{\Theta^2}(t). \tag{11.6}$$

The input power per unit area (divide by the area of two footpoints!), using equation (11.5) for $\overline{\Theta^2}(t)$, is then

$$P = \frac{3R^2 B_z^2 \tau_c \overline{V^2}}{16\pi\lambda^2 L}. \tag{11.7}$$

Let $B_z = 100$ G, $L = 10^5$ km, and $R = \lambda$ for an active region corona. With these values the heating rate is $P = 5 \times 10^5\,\mathrm{ergs\,cm^{-2}\,sec^{-1}}$, one order of magnitude lower than the 10^7 required for active region heating (Withbroe and Noyes 1977). Note the sharp dependence on $(R/\lambda)^2$, however.

For the quiet sun, the model works somewhat better. Here the requirement is approximately $P = 3 \times 10^5$. There are fewer photospheric flux elements, so that we may take $B_z = 10$ G and $R = 3\lambda$. For these values the random twisting of individual tubes gives a heating rate one sixth that required.



11.4 Braided coronal tubes. For simplicity the tubes lie between two planes rather than forming an arch.

11.4 The Parker model

The previous section treated the effect of random motions on a single flux tube. For regions where the footpoints are bunched close together, several tubes can become braided. This braiding greatly enhances the efficiency of stochastic heating. Consider a magnetic field residing between the planes $z = 0$ and $z = L$. These boundary planes correspond to the photosphere; the corona lies between them (see figure 11.4). We allow the photospheric fields to be localized into discrete footpoints; in thin transition regions $0 < z < \epsilon$ and $L - \epsilon < z < L$ the flux tubes expand until they touch each other. The transition regions will be ignored in our discussion. Between $z = \epsilon$ and $z = L - \epsilon$ the axial field will be assumed to be uniform, $B_z = const.$ Start at $t = 0$ with a purely axial field, $\vec{B}_\perp = 0$. The endpoints of the field lines then random walk about each other until they are thoroughly tangled. This generates a nonzero $\vec{B}_\perp$. What is the equilibrium energy of a tangled magnetic field?

First we define two lengths. Let $\ell(t)$ be the total transverse length of a line (project the line into the $x - y$ plane and take the total arc length). Also let $s(t)$ be the transverse distance between the top and bottom endpoints of the field line. Unless the field line is absolutely straight, $\ell(t) > s(t)$. In any case, we may estimate $|\vec{B}_\perp|$ by

$$|\vec{B}_\perp|/\ell = B_z/L. \tag{11.8}$$

To aid our understanding of this problem, we first discuss a simple case *without* tangling. Here the field consists of discrete flux tubes whose footpoints are spaced far apart from each other. If $s(t)$ is much less than the distance d between footpoints, then each fluxtube can

evolve more or less independently. To a rough approximation, the field lines will all be straight, with $\ell(t) = s(t)$. The average excess energy (over the trivial field $\vec{B}_\perp = 0$) per unit surface area is thus

$$\overline{B_\perp^2} L/16\pi = \overline{s^2}(t) B_z^2/16\pi L. \tag{11.9}$$

For random motions with diffusion coefficient K, the mean square distance travelled (by two endpoints) is $\overline{s^2}(t) = 8Kt$. Equation (11.9) thus gives an input power

$$P = KB_z^2/2\pi L. \tag{11.10}$$

For the active region parameters used in the previous section, this give a heating power of $P = 3\times10^5$ ergs cm^{-2} sec^{-1}, somewhat smaller than the Sturrock-Uchida value.

Now suppose that the footpoints are close enough to each other so that their coronal flux tubes become braided about each other, as in figure 11.4. Because of the tangling the field lines can no longer straighten out, and $\ell(t) > s(t)$. Parker (1983) considered the extreme case where $\ell(t)$ equals the total distance travelled by the two endpoints, $\ell(t) = 2Vt$ ($V \equiv (\overline{V^2})^{1/2}$). In this case

$$P = V^2 B_z^2 t/2\pi L. \tag{11.11}$$

Note the dependence on time t here. In the previous models a particular photospheric motion is equally likely to increase or decrease $s(t)$ or $\Theta(t)$. The square of these quantities grows linearly with time; but $d\overline{s^2}(t)/dt$ and $d\overline{\Theta^2}(t)/dt$ are independent of time. For braiding, however, $\ell^2(t)$ grows quadratically in time.

Berger (1990) has examined the braiding process more closely. In an analysis of braids involving three flux tubes, a topological invariant $\mathcal{T}$ (the tangle number) was introduced to measure the amount of braiding. (The braid in figure 11.1 has $\mathcal{T} = 14$, while figure 11.4 has $\mathcal{T} = 5$.) This number increases linearly with time. Also the length ℓ goes as the number of 'tangles' $\mathcal{T}$ times the distance travelled per tangle (about the diameter D of a flux tube). So we can write $\ell(t) = \gamma D\mathcal{T}$ where $\gamma \approx 1$. If the distance d between photospheric footpoints is large compared to the coherence length λ of the motion, then the footpoints must random walk before one footpoint moves around another footpoint and hence increases the tangle number by 1. This implies $\ell(t)$ should be multiplied by the factor $D\lambda/d^2$. Furthermore, there is indeed some probability that $\mathcal{T}$ can decrease as well as increase as the motion proceeds. The efficiency of tangling is reduced by some factor ϵ due to the possibility of untangling. With

these considerations $\ell(t)$ changes to

$$\ell(t) = \epsilon\gamma(D\lambda/d^2)2Vt. \tag{11.12}$$

The power can be computed if we make the following assumption: reconnection will not balance the input of magnetic energy until the transverse field has reached some significant level. Let us write

$$B_\perp/B_z = \ell/L = \mu. \tag{11.13}$$

Note that neighbouring flux tubes in a braided field are misaligned with respect to each other. The angle between neighboring tubes is roughly $2\tan^{-1}\mu$. If we suppose that this angle must be 30° before reconnection is triggered, then $\mu = 0.27$. This tells us t: energy balance is reached when $\ell(t) = L\mu$. The power output becomes

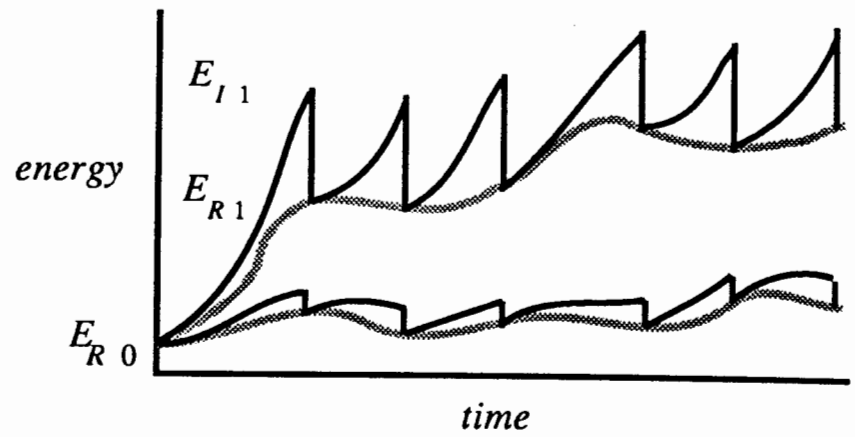
$$P = \epsilon\gamma\mu(D\lambda/d^2)VB_z^2/4\pi. \tag{11.14}$$

A numerical simulation for three tubes yielded $\epsilon = 1/4$. This number may go up somewhat for more than three tubes. For the active region values we have been using, with $\lambda \approx d$, we obtain $P = 0.5 \times 10^7\,\mathrm{ergs\,cm^{-2}\,sec^{-1}}$. This estimate makes the Parker model a viable model for active region heating. The time it takes to build up to the equilibrium amount of tangling ($\mathcal{T} \approx 13$) is $t = 5 \times 10^4$ sec.

11.5 The Heyvaerts-Priest model

The Heyvaerts-Priest model (Heyvaerts and Priest 1985; Browning and Priest 1986) assumes that turbulent reconnection processes episodically dissipate free magnetic energy in the corona. A turbulent resistive plasma is extremely difficult to model, so the field evolution is separated into an ideal phase and a relaxation phase. During the ideal phase (lasting a time τ_i) resistivity is ignored; the field evolves through a sequence of equilibria in response to surface motions. During the short relaxation phase (duration $\tau_r << \tau_i$) reconnection converts magnetic energy into heat. The cycle then repeats with a new ideal phase. Let us call the magnetic energy at the end of the ideal phase E_{Ii} where $i = 1, 2, \ldots$ labels the cycle. The energy after relaxation is E_{Ri}. Starting with E_{R0} at $t = 0$ the energy evolves through a sequence $E_{R0}, E_{I1}, E_{R1}, E_{I2}, \ldots$. The free energy released by relaxation after cycle i is $E_{Ii} - E_{Ri}$ (see figure 11.5).

If there were no constraints on the field evolution during the resistive phase, the endstate would be a vacuum field consistent with $\vec{B}\cdot\hat{n}$ at the surface. One would then obtain the values of E_{Ri} by calculating vacuum field energies. To obtain E_{Ii}, one must take the



11.5 The dotted curves show minimum energy vs. time. Lower curve – without H conservation; upper curve – with H conservation. The solid curves show the evolution of a field undergoing episodic reconnection.

(known) field configuration after the $(i-1)^{th}$ relaxation phase, and evolve it according to the surface motions. This is a nonlinear calculation, but can be done if τ_i is not too large and if the fields and motions possess some symmetry.

Unfortunately, for the case of relaxation to vacuum fields, the average heating rate $(E_{Ii} - E_{Ri})/(\tau_i + \tau_r)$ is rather low. The heating rate can be considerably enhanced, however, if we take into account magnetic helicity conservation.

Magnetic helicity measures the net twisting and shearing of field lines (Berger 1988). For inhomogeneous boundary conditions ($\vec{B}\cdot\hat{n} \neq 0$) the helicity can be written (Berger and Field 1984; Finn and Antonsen 1985)

$$H = \int(\vec{A} + \vec{A}_0)\cdot(\vec{B} - \vec{B}_0)\,d^3x \tag{11.15}$$

where $\vec{B}_0$ is the vacuum field satisfying $\vec{B}_0\cdot\hat{n} = \vec{B}\cdot\hat{n}$ and $\nabla\times\vec{A} = \nabla\times\vec{B}$, $\nabla\times\vec{A}_0 = \nabla\times\vec{B}_0$.

Helicity is absolutely conserved in the absence of resistivity. It can be transferred from the solar interior to the corona if there are surface motions in regions where $\vec{B}\cdot\hat{n} \neq 0$. Furthermore, strict limits an be placed on helicity nonconservation during reconnection (Berger 1984). These limits are entirely independent of the reconnection geometry; they depend only on η, E_{Ii}, E_{Ri}, and τ_r: in particular the net helicity lost during a relaxation phase $H_{Ii} - H_{Ri}$ satisfies

$$(H_{Ii} - H_{Ri})^2 \leq 64\pi^2(E_{Ii}^2 - E_{Ri}^2)\eta\tau_r. \tag{11.16}$$

Helicity has units of magnetic flux squared. If Φ is the axial flux in a coronal loop, then H/Φ^2 measures the average twist of field lines about the axis of the loop ($H/\Phi^2 = 1$ corresponds to one net turn through an angle 2π). For a loop of radius 10^4 km, length 10^5 km, $H/\Phi^2 = 1$, and $\tau_r = 10^3$ sec, one estimates (Berger 1984)

$$(H_{Ii} - H_{Ri})/H_{Ii} < 10^{-5}. \qquad (11.17)$$

The minimum energy state for a given helicity is a constant α force-free field, $\nabla \times \vec{B} = \alpha \vec{B}$ with $\nabla\alpha = 0$. The Heyvaerts-Priest model assumes Taylor relaxation (Taylor 1974), where relaxation proceeds all the way the to the constant α state. Constant α fields thus replace vacuum fields in the calculation of $E_{Ii} - E_{Ri}$.

The relaxation phase, constrained now by helicity conservation, is less effective in removing magnetic energy – E_{Ri} is greater for a constant-α field than for a vacuum field. At first sight this should reduce the heating rate. However, E_{Ii} also changes when helicity conservation is included. The nonlinear force free field attained at the end of the ideal phase depends not only on the surface motions during the ideal phase, but also on the initial field configuration. But this intial configuration is just the state reached by the previous relaxation phase. If this state is a constant-α field with currents, the nonlinear evolution will amplify those currents and yield a higher E_{Ii}. For coherent photospheric motions, for example the steady shearing of an arcade, the model can produce a heating rate consistent with observations.

In general, the volume over which relaxation occurs affects the heating rate. The twist parameter H/Φ^2 can be high for a single loop or arcade, but will be lower for larger structures. Thus a localized relaxation process involving a few coronal loops will yield a higher heating rate than if relaxation involves many loops. Furthermore, If τ_i is much greater than the coherence time of the photospheric motion, then a statistical treatment must be used. Random photospheric motions of several footpoints yield a low root mean square H/Φ^2, drastically reducing the efficiency of the heating process (Choudhuri 1986).

One direction for further work involves the problem of incomplete relaxation. There is no *a priori* reason to suppose that Taylor relaxation will happen in the corona; the field may not relax all the way down to the constant-α state. For example, other constraints besides total helicity conservation may play a role during relaxation.

11.6 Dissipation of Coronal Magnetic Fields

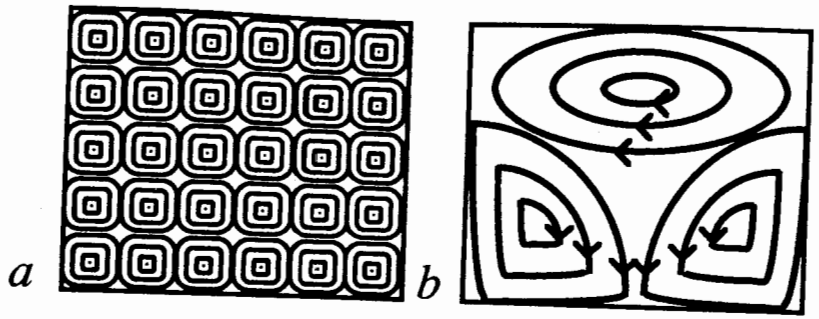
Motions in the photosphere and convection zone continually add to the complexity of the coronal field. This cannot go on forever; eventually some dissipation process must balance the input of structure. Note that this difficulty exists even if magnetic energy dissipation is not the primary source of heating.

Consider a coronal arch as shown in figures 11.1 and 11.4. We will look at $\vec{B}_\perp$, the field transverse to the axis of the arch. The axial field B_z itself is little affected by dissipation, since resistivity conserves the total axial flux. The energy *input* processes discussed in §11.3 and §11.4 have timescales τ_i of hours or days. The diffusion timescale for $\vec{B}_\perp$ can be written $\ell_\perp^2/\eta$, where $\ell_\perp$ measures the distance over which $\vec{B}_\perp$ changes direction, $\ell_\perp \sim |\vec{B}_\perp|/|\nabla \times \vec{B}_\perp|$. Consider as an example a leisurely input rate with $\tau_i = 1$ week $= 6 \times 10^6$ sec. Balancing input and output leads to $\ell_\perp \approx 2$ km. The photospheric lengthscale d_g introduced in §11.1 is 2–3 order of magnitudes greater. How can the smaller lengthscale $\ell_\perp$ be generated?

Parker (1972, 1979, 1987) proposes that most dissipation occurs at thin current sheets, where $\ell_\perp$ is very small. He suggests that these thin current sheets must be present in the coronal field. The reason is that for most field topologies no smooth equilibria exist without current singularities. van Ballegooijen (1986) has suggested an alternative to current sheet formation: a cascade of magnetic energy to very small scales leads to the small $\ell_\perp$ and high current densities needed for dissipation.

However, even if current sheets exist, we still need to know how widespread they are. If the number of current sheets is insufficient, coronal heating must then rely on the dissipation of other structures such as fine-scale current filaments. Here we will show that this cannot happen – almost all of the current in a tangled field must indeed reside in thin current sheets.

Figure 11.6 demonstrates the problem. Does the transverse field consist of numerous small-scale islands, as in figure 11.6a, or relatively few islands separated by narrow current layers, as in 11.6b? Now, the circular lines in figure 11.6a are actually the projections of spiral field lines. In three dimensions the field lines in each island of concentric circles spiral about a central axis; this axis crosses the plane of the figure at the central point of the island. How many spirals does a field line have? We will use this question to show that the field geometry of figure 11.6a is not possible.



11.6 Lines of the transverse field in a coronal arch. Note that there must be a current sheet between the left and upper islands in 11.6b.

Let Θ give the net angle through which a field line in an island spirals about the central axis. For a spiral a small distance r away from the axis,

$$|\vec{B}_\perp|/r\Theta = B_z/L. \tag{11.18}$$

We model figure 11.6a by the field

$$B_x = a \sin kx \cos ky; \quad B_y = -a \cos kx \sin ky \tag{11.19}$$

where $k = \pi/2\ell_\perp$ and $\ell_\perp$ is the radius of an island. By Stokes' theorem

$$|\vec{B}_\perp| = r|(\nabla \times \vec{B}_\perp)_z|/2. \tag{11.20}$$

For our model field, this leads to

$$\Theta = \pi L/4B_z \frac{a}{\ell_\perp}. \tag{11.21}$$

In order to evaluate this equation for Θ, we must find $a/\ell_\perp$. This can be done by balancing energy input from the photosphere with Ohmic dissipation. Let us assume the *least* efficient input mechanism – the simple random walk model presented at the beginning of §11.4. The energy input per unit area is given by equation (11.10). A simple calculation yields a balance of dissipation and input (per unit photospheric area) when

$$\frac{\pi\eta a^2 L}{32\ell_\perp^2} = \frac{KB_z^2}{2\pi L}, \tag{11.22}$$

or

$$\frac{a}{\ell_\perp} = \frac{4B_z}{\pi L}\sqrt{K/\eta}. \tag{11.23}$$

Going back to equation (11.21) for Θ, we obtain the simple result

$$\Theta = \sqrt{K/\eta}. \tag{11.24}$$

For $\eta = 10^{-6}\,\mathrm{km^2\,sec^{-1}}$ and $K = 200\,\mathrm{km^2\,sec^{-1}}$, $\Theta = 1.4 \times 10^4$ radians; i.e. a field line spirals about the central axis through some 2,000 complete turns! Note that Θ is an ideal MHD invariant. We expect a root mean square Θ equivalent to only a few turns (see §11.3). The immense value needed for significant Ohmic dissipation is thus not reasonable. This argument rules out dissipation in small magnetic islands. Instead, dissipation must predominantly occur in a small fraction of the volume, near current sheets as in figure 11.6b.

References

Berger MA and Field GB (1984) The topological properties of magnetic helicity *J. Fluid Mechanics* **147** *133.*

Berger MA (1984) Rigorous new limits on magnetic helicity dissipation in the solar corona *Geophysical and Astrophysical Fluid Dynamics* **30** *79.*

Berger MA (1988) An energy formula for nonlinear force-free magnetic fields *Astronomy and Astrophysics* **201** *355.*

Berger MA (1990) Braided magnetic flux ropes and coronal heating in "Magnetic Flux Ropes", Bermuda Chapman conf. proceedings (C. Russell, E. Priest, and L.C. Lee, eds.) p 251

Browning PK and Priest ER (1986) *Astronomy and Astrophysics* **159** *129.*

Choudhuri AR (1986) in "Coronal and Prominence Plasmas" (AI Poland, ed.) (NASA CP 2442) p 451.

Finn JH and Antonsen TM (1985) Magnetic Helicity: What is it and What is it Good For? *Comments on Plasma Physics and Controlled Fusion* **9***111*

Frazer, EN and Stenflo JO (1972) On the small scale structure of solar magnetic fields *Solar Physics* **27** *330*

Heyvaerts J and Priest ER (1985) Coronal Heating by Reconnection in D.C. Current Systems: A Theory Based on Taylor's Theorem *Astronomy and Astrophysics* **137** *63.*

Lou YQ and Rosner R (1986) The damping of the Alfvén mode in stochastic magnetic fields *Astrophysical J.* **309** *874.*

Parker EN (1972) Topological dissipation and the small-scale fields in turbulent gases *Astrophysical J.* **174** *499*

Parker EN (1979) *Cosmical Magnetic Fields* Oxford, U.K.

Parker EN (1983) Magnetic Neutral Sheets in Evolving Fields, II. Formation of the Solar Corona *Astrophysical J.* **264** *642.*

Parker EN (1987) Magnetic reorientation and the spontaneous formation of tangential discontinuities in deformed magnetic fields *Astrophysical J.* **318** *876.*

Priest ER (1982) *Solar Magnetohydrodynamics* (Dordrecht: Reidel)

Similon PL and Sudan RN (1989) Energy dissipation of Alfvén wave packets derformed by irregular magnetic fields *Astrophysical J.* **336** *442.*

Smithson RC (1973) Videomagnetograph studies of solar magnetic fields I. Magnetic field diffusion in weak plage regions *Solar Physics* **29** *365.*
Song Y and Lysak RL (1989) Evaluation of twist helicity of FTE flux tube *J. Geophys. Res.* **94** *5273.*
Sturrock PA and Uchida Y (1981) Coronal heating by stochastic magnetic pumping *Astrophysical J.* **246** *331.*
Taylor JB (1974) Relaxation of Toroidal Plasma and Generation of Reverse Magnetic Fields *Physical Review Letters* **33** *1139.*
van Ballegooijen AA (1986) *Astrophysical J.* **311***1001.*
Velli M and Hood AW (1989) Resistive tearing in line-tied magnetic fields: slab geometry *Solar Physics* **119** *107.*
Withbroe GL and Noyes RW (1977) Mass and energy flow in the solar chromosphere and corona *Annual Review of Astronomy and Astrophysics* **15** *363.*
Wright AN and Berger MA (1989) The effect of reconnection upon the linkage and interior structure of magnetic flux tubes *J. Geophys. Res.* **94** *1295.*

12

Hydromagnetic Equilibrium

K. TSINGANOS

University of Crete and Research Center of Crete, GREECE

Abstract

The methodology and state of the art of constructing analytical MHD equilibria is reviewed. The formalism is illustrated by two classes of analytical solutions. In the first, a typical example of MHD equilibrium in the uniform gravitational field of a polytropic atmosphere is constructed. The solution extends the familiar Kippenhahn–Schlüter model for a quiescent solar prominence by including flows along the magnetic field lines. In the second, a typical example of MHD equilibrium in the nonpolytropic stellar atmosphere surrounding the central gravitational field of a star is constructed. The solution extends the familiar Parker model of the solar/stellar wind by including a magnetic field without spherical symmetry and hydromagnetic critical points in the rotating outflow.

12.1. MHD Equilibrium

Ideal MHD is the most basic single-fluid model for determining the macroscopic equilibrium and stability properties of a plasma where the Debye length, l_D, the cyclotron radius r_c and the mean free path, λ, are small compared to the characteristic scale of the system, $r >> \lambda >> r_c >> l_D$ (Parker 1979, Priest 1984). This is the case in many astrophysical phenomena, like the various hydromagnetic structures of the solar/stellar atmosphere where, in addition, the bulk flow speed is usually nonrelativistic and the electrical conductivity of the gas is extremely high, while the viscosity is negligibly low. In those circumstances one uses the following set of equations to describe the steady interaction of the magnetofluids in the ambient gravitational field,

$$\nabla \cdot \mathbf{B} = 0, \qquad \nabla \cdot (\rho \mathbf{V}) = 0\,, \tag{12.1a}$$

$$\nabla \times (\mathbf{V} \times \mathbf{B}) = 0\,, \tag{12.1b}$$

$$\rho(\mathbf{V}\cdot\nabla)\mathbf{V} = -\nabla P + \frac{1}{4\pi}(\nabla \times \mathbf{B}) \times \mathbf{B} - \rho\nabla g\,, \tag{12.1c}$$

$$T = \frac{m_p}{2k}\frac{P}{\rho}\,, \tag{12.1d}$$

for the bulk flow speed $\mathbf{V}(x_1, x_2, x_3)$, the magnetic field $\mathbf{B}(x_1, x_2, x_3)$, the density $\rho(x_1, x_2, x_3)$, the pressure $P(x_1, x_2, x_3)$ and temperature $T(x_1, x_2, x_3)$ in the orthogonal coordinates (x_1, x_2, x_3), with $g(x_1, x_2, x_3)$ the gravitational potential, m_p the proton mass and k Boltzmann's constant.

The system of equations (12.1) needs to be closed with an equation for the energy transport. However, the present unpleasant reality is that a detailed equation of state $P = P(\rho, S)$, where S is the entropy, is completely unknown due to poor knowledge of the various energy transport and deposit mechanisms operating along the flow. The system (12.1) may instead be closed via a polytropic relationship between pressure and density, with a constant polytropic index γ, $1 < \gamma < 5/3$, (Parker 1963), i.e.,

$$\gamma \equiv \frac{d\ln P}{d\ln\rho} = const. \tag{12.2a}$$

This means that we effectively specify *a priori* the distribution of the heating/cooling mechanisms in the flow in such a way as to correspond to the constant value of γ in (12.2a). Since this *a priori* specification of a constant γ may correspond to quite an artificial heating/cooling distribution along the flow, as well as streamline shape, the opposite approach may be interesting, i.e., to specify the shape of the streamlines, as can be inferred from observations. Then we can deduce from the following energy conservation equation the heating/cooling required to support such a flow pattern (Low and Tsinganos 1986, Tsinganos and Low 1989, Tsinganos and Trussoni 1990, 1991):

$$3(\frac{k\rho}{m_p})(\mathbf{V}\cdot\nabla)T - 2(\frac{k\rho}{m_p})(\mathbf{V}\cdot\nabla)\rho = \rho\sigma\,. \tag{12.2b}$$

In the above expression $\rho\sigma$ denotes the rate of energy deposition per unit volume of the fluid. The resulting value of the effective variable γ,

$$\gamma \equiv \frac{\partial\ln P}{\partial\ln\rho}\Big|_{streamline}\,, \tag{12.2c}$$

may then be compared to some characteristic values, such as $\gamma = 1$ (isothermal atmosphere), γ=3/2 (Parker polytrope) or γ=5/3 (adiabatic expansion.

Finally, in some cases it is convenient and physically interesting to close the system (12.1) via the assumption ρ = const., or, $\rho = \rho(A)$ (Tsinganos 1981, de Ville and Priest 1991), or, by prescribing the temperature distribution along each streamline A=*const.* in a physically meaningful way (Low 1975, Tsinganos 1982).

In the following we shall integrate one by one the conservation laws expressed by Eqs.(12.1) together with one of equations (12.2), in the well known manner (e.g., Tsinganos 1982). This integration of the partial differential equations results in several arbitrary functions. What is not well known, however, is how to choose selfconsistently these arbitrary functions such that, first, this choice has a physical basis, and second, the resulting differential equations may be solved exactly.

12.2. Symmetric Equilibria

Nonsingular 3-D MHD equilibria that extend in all space are not known to exist (Parker 1979, Tsinganos, Distler and Rosner 1984). The 2-D character of MHD equilibria may be seen from the requirement that magnetic fieldlines together with streamlines are constrained to lie on 2-D surfaces,

$$\mathbf{V}\times\mathbf{B} = \nabla\Phi, \qquad \mathbf{V}\cdot\nabla\Phi = \mathbf{B}\cdot\nabla\Phi = 0\,, \tag{12.3}$$

Consider then the integrals of the steady MHD equations (12.1) in an orthogonal system (x_1, x_2, x_3) with line elements $h_i(x_1,x_2)$, $i = 1, 2, 3$, wherein the coordinate x_3 is ignorable and the system is correspondingly symmetric (Tsinganos 1982),

$$\mathbf{B}(x_1, x_2) = \nabla A \times \nabla x_3 + \nabla x_3 \frac{h_3^2\Phi_A\Psi_A - L(A)\Psi_A}{1 - \Psi_A^2/4\pi\rho}\,, \tag{12.4a}$$

$$\mathbf{V}(x_1, x_2) = \frac{\Psi_A}{4\pi\rho}\nabla A \times \nabla x_3 + \nabla x_3 \frac{h_3^2\Phi_A - L(A)\Psi_A^2/4\pi\rho}{1 - \Psi_A^2/4\pi\rho}\,, \tag{12.4b}$$

$$\left[A\,,\ \frac{\Delta A - \Psi_A\Delta\Psi}{4\pi\rho}\right] = \left[h_3^2V_3^2\,,\ \frac{1}{2h_3^2}\right] + \left[h_3^2B_3^2\,,\ \frac{1}{8\pi\rho h_3^2}\right] + \left[\frac{1}{\rho}\,,\ P\right] \tag{12.4c}$$

where $h_3\Delta A = -\mathbf{e_3}\cdot\nabla\times\mathbf{B}$, $h_3\Delta\Psi = -\mathbf{e_3}\cdot\nabla\times\mathbf{V}$ and the notation of

Poisson brackets is used. The integrals $\Psi(A)$ [streamfunction], $\Phi(A)$ [induction potential function] and $L(A)$ [total angular momentum function] are arbitrary apart from some relationships among them that arise at the fixed points of the differential equations (12.1). Noting that Eq. (12.4c) relates the magnetic flux function $A(x_1, x_2)$ and the density $\rho(x_1, x_2)$, the second needed relationship between A and ρ may be obtained from expressions (12.2). Thus, in the polytropic case, Eq. (12.2a), the last term in Eq. (12.4c) vanishes and a fourth integral arises from the conservation of the total energy flux density per unit of mass flux density, $F(A)$,

$$F(A) = \int \frac{dP}{\rho} + \frac{\mathbf{V}^2}{2} + g - \frac{\Phi_A}{\Psi_A} h_3 B_3 \,. \tag{12.5a}$$

In the nonpolytropic case, Eq. (12.2b), we are free to impose an extra relation, for example, the shape of the streamlines, or some other relationship between the components of the pressure, P_o and P_1,

$$P(R, A) = P_o(R) + P_1(R)A \,. \tag{12.5b}$$

Finally, in the case where the density is constant along each streamline, $\rho = \rho(A)$, the fourth integral is the generalized Bernoulli integral $\Pi(A)$,

$$\Pi(A) = P + \frac{\rho \mathbf{V}^2}{2} + \rho g - \frac{\Phi_A}{\Psi_A} \rho h_3 B_3 \,, \tag{12.5c}$$

while if the temperature distribution is given, we have

$$\begin{aligned} P(A, Z) &= P_o(A) \exp\left[- \int_{z_o}^{z} \frac{m dZ'}{kT(A, Z')} \right], \\ Z &= \frac{V^2}{2} + g - \frac{\Phi_A}{\Psi_A} h_3 B_3 \,, \end{aligned} \tag{12.5d}$$

i.e., that $P_o(A)$ is the fourth integral of the system.

In the following sections we present two classes of analytical solutions that illustrate the use of the previous formalism. In section 3 a typical example of MHD equilibrium in the *uniform gravitational field of a polytropic atmosphere* is constructed, while in section 4 a typical example of MHD equilibrium in the *nonpolytropic stellar atmosphere* surrounding the *central gravitational field* of a star is constructed.

12.3. MHD Equilibrium in Uniform Gravity

As the first example consider the problem of the hydromagnetic equilibrium of plasma in a uniform gravitational field, a situation pertinent to quiescent solar prominences. This solution may be regarded as an extension of the classical Kippenhahn-Schlüter (1957) model for a quiescent prominence [see Priest (1984) for a comprehensive review].

12.3.1 Magnetohydrostatic Equilibrium [$\mathbf{V} = 0$, Kippenhahn-Schlüter Solution]

Consider the simplest choice of the integrals of the MHD equations (12.1) combined with a polytropic relationship like Eq. (12.2a),

$$\begin{aligned} \Psi(A) = \Psi_o, \quad \Phi(A) = \Phi_o, \quad \Omega(A) = -L(A)\Psi(A) = \Omega_o, \\ F(A) = F_o + (g/B_o)A \,, \end{aligned} \tag{12.6a}$$

where Ψ_o, Φ_o, Ω_o and F_o are constants. In addition we make the *ansätz* that the magnetic flux function A has the following form,

$$A(y, z) = B_o z - \int B_z(y) dy \,, \tag{12.6b}$$

such that the magnetic field is,

$$\mathbf{B} = [\Omega_o \,, B_o \,, B_z(y)] \,, \tag{12.6c}$$

in an orthogonal coordinate system xyz where the z-axis points in the opposite direction to the uniform gravitational field g. The three components of the momentum balance equation are then,

$$\frac{\mathrm{d}}{\mathrm{d}y} [B_y B_x] = 0 \,, \tag{12.7a}$$

$$\frac{\mathrm{d}}{\mathrm{d}y} \left[P + \frac{B_z^2}{8\pi} \right] = 0 \,, \tag{12.7b}$$

$$\frac{\mathrm{d}}{\mathrm{d}y} \left[\frac{B_y B_z}{8\pi} \right] = \rho g \,. \tag{12.7c}$$

For an isothermal atmosphere, $\mathrm{T} = T_o$, $V_s^2 = 2kT_o/m_p$, and Eq. (12.7b) yields

$$V_s^2(\rho_o - \rho) = \frac{B_z^2}{8\pi} \,, \tag{12.8}$$

where ρ_o is a constant. It is evident that ρ_o is the maximum density which is allowed in the atmosphere. This density is attained when $B_z = 0$, say, at $\mathrm{y} = 0$. In dimensionless units, $\mathrm{Y} = \mathrm{y/L}$ ($\mathrm{L} = V_s^2/g$

is the scale height), $R = \rho/\rho_o$ and $H = B_z/B_o$, the system of Eqs. (12.7) takes the form

$$B_x = const., \tag{12.9a}$$

$$H = \pm\sqrt{\beta(1-R)}, \tag{12.9b}$$

$$\frac{dH}{dY} = \frac{\beta R}{2}. \tag{12.9c}$$

where β is the familiar plasma ratio at the origin, $\beta = 8\pi V_s^2\rho_o/B_o^2$. Substituting Eq. (12.9b) in Eq. (12.9c) we obtain

$$\pm dY = \frac{dR}{R\sqrt{\beta(1-R)}}. \tag{12.9d}$$

Finally, integrating Eq. (12.9d) and then comparing the result with Eq. (12.9b) we obtain

$$R(Y) = \frac{1}{\cosh^2(\sqrt{\beta}\,Y/2)}, \tag{12.10a}$$

$$H(Y) = \pm\sqrt{\beta}\tanh(\sqrt{\beta}\,Y/2). \tag{12.10b}$$

The dimensionless density R and magnetic field H are expressed in terms of a single parameter: the plasma ratio β at the origin, Y = 0, where the density is maximum and the magnetic field lines are horizontal, $B_z = 0$. Notice that for given maximum density ρ_o the z-component of the magnetic field which supports the plasma weight against gravity is asymptotically

$$\frac{B_z(\infty)}{B_o} = \pm\sqrt{\beta} = \pm\sqrt{8\pi\rho_o V_s^2/B_o^2}. \tag{12.11}$$

The minimum radius of curvature r_o of the field lines at the point of maximum density is then

$$r_o = \frac{V_A^2}{g}. \tag{12.12}$$

Notice also that in this solution the total pressure $[P + B^2/8\pi]$ is constant everywhere, with the result that the plasma weight is supported against gravity by the tension of the magnetic lines alone. The density is maximum at the origin where the field lines have the greatest curvature and tension. The density tends asymptotically to zero at infinity, where the field lines become straight lacking any curvature and tension. The solution (12.10) is well known as the Kippenhahn-Schlüter model for a quiescent prominence (Kippenhahn-Schlüter 1957, Low 1975). Note that most of the plasma is confined horizontally within a few scale heights; and the extent of the volume containing most of the plasma is smaller the smaller is β.

In the above Kippenhahn-Schlüter solution the atmosphere is compressible only in the horizontal direction y. It is an interesting fact that this solution can exist in a plane-parallel stratified atmosphere as well (Low, Hundhausen and Zweibel 1983). In that case the total density ρ(y,z) is given by the expression:

$$\rho(y,z) = \rho_{\|}e^{(-z/L)} + \frac{\rho_0}{\cosh^2(\sqrt{\beta}\,Y/2)}, \tag{12.10c}$$

where $\rho_{\|}$ is the density at z=0, $y \to \pm\infty$.

*12.3.2 Hydrostatic Equilibrium, [**B** = 0]*

It is instructive to consider for a moment the steady state of the similar, one-dimensional velocity field

$$\rho\mathbf{V} = [\rho(y)V_x(y), \rho(y)V_y(y), \rho(y)V_z(y)], \tag{12.13}$$

in the same isothermal atmosphere. Then the continuity equation is satisfied identically if $\rho(y)V_y(y) = const$, while the remaining force balance equations read,

$$\frac{d}{dy}[V_y] = 0, \tag{12.14a}$$

$$\frac{d}{dy}\left[P + \rho V_y^2\right] = 0, \tag{12.14b}$$

$$\frac{d}{dy}[\rho V_y V_z] = \rho g. \tag{12.14c}$$

Noting the analogy of Equations (12.14) to Equations (12.7) of the previous case of magnetostatic equilibrium, we can integrate Equation (12.14b) to obtain ρ=constant. Notice that this result is drastically different from that expressed in Equation (12.10a): the streamlines of the flow are the familiar parabolic paths of a projectile fired against the uniform gravitational field instead of the sagged, under the plasma weight, magnetic field lines. None of the interesting features found in the one-dimensional magnetostatic equilibrium is found in the case of the one-dimensional hydrostatic equilibrium. This physical difference is better understood when we compare the Maxwell and Reynolds stress tensors (Parker 1979),

$$M_{ij} = -\left[P + \frac{\mathbf{B}^2}{8\pi}\right]\delta_{ij} + \frac{B_iB_j}{4\pi}, \tag{12.15a}$$

$$R_{ij} = -P\delta_{ij} - \rho V_iV_j. \tag{12.15b}$$

The non-uniform part of the Maxwell's stress tensor represents tension along a magnetic line while the non-uniform part of the Reynolds stress tensor represents compression along a stream line. The inevitable result is that magnetic lines are stretched tight whereas stream lines tend to buckle. It becomes of interest then to know which of the two stresses dominates when we have flows along the magnetic field lines, a goal that we pursue in the following.

12.3.3 Hydromagnetic Equilibrium in Uniform Gravity [$\mathbf{B} \neq 0$, $\mathbf{V} \neq 0$]

Having discussed separately the simplest cases of pure magnetostatic and hydrostatic equilibrium in uniform gravity in the previous sections, we proceed now with a step-by-step construction of an analytical solution for the composite case of dynamical equilibrium ($\mathbf{V} \neq 0$). It will be seen that this general case of MHD equilibrium is much more complex than the simpler pure hydrostatic and magnetostatic cases discussed above.

12.3.3.1 The Solution

Take again the z-axis of the orthogonal system xyz to point in the opposite direction to the uniform acceleration of gravity g. Use the previous formulation (12.1) with the following choice of the free integrals Ψ, Φ, L and F,

$$\Psi(A) = \Psi_o + 4\pi k A, \quad \Phi(A) = \Phi_o,$$
$$L(A) = 0, \quad F(A) = F_o + (g/B_o)A, \tag{12.16a}$$

where Ψ_o, Φ_o, F_o, B_o and k are constants. With the *ansätz*

$$A(y,z) = B_o z - \int B_z(y)dy\,, \tag{12.16b}$$

the following set of solutions is obtained,

$$\mathbf{B} = [0\,, B_o\,, B_z(y)]\ , \qquad \rho\mathbf{V} = k\mathbf{B}\,, \tag{12.17}$$

$$\frac{\mathrm{d}}{\mathrm{d}y}\left[P + \frac{B_z^2}{8\pi} + \frac{k^2 B_o^2}{\rho}\right] = 0\,, \tag{12.18a}$$

$$\frac{\mathrm{d}}{\mathrm{d}y}\left[\frac{B_o B_z}{4\pi}\left(1 - \frac{4\pi k^2}{\rho}\right)\right] = \rho g\,. \tag{12.18b}$$

Assuming for simplicity an isothermal atmosphere, $P = V_s^2\rho$, Eq.

(12.18a) can be integrated to give,

$$\rho V_s^2 + \frac{B_z^2}{8\pi} + \frac{k^2 B_o^2}{\rho} = \rho_1 V_s^2\,, \tag{12.19}$$

where ρ_1 is a positive constant. Combining Eqs. (12.18b) and (12.19) we obtain

$$B_z(\rho) = \pm\sqrt{\frac{8\pi V_s^2}{\rho}(-\rho^2 + \rho\rho_1 - \gamma^2)}\,, \tag{12.20a}$$

$$\frac{\mathrm{d}}{\mathrm{d}\rho}\left[\frac{B_o B_z(\rho)}{4\pi}\left(1 - \frac{4\pi k^2}{\rho}\right)\right] = \rho g \frac{\mathrm{d}y}{\mathrm{d}\rho}\,. \tag{12.20b}$$

where $\gamma = kB_o/V_s$. Eq. (12.20b) - with $B_z(\rho)$ substituted from Eq. (12.20a) - is then the final equation which gives the density ρ as a function of the horizontal distance y. Substituting Eq. (12.20a) in Eq. (12.20b) we obtain the following first-order ordinary differential equation,

$$\pm\sqrt{\frac{8\pi V_s^2}{B_o^2}}\mathrm{d}Y = \left[\frac{C(\rho)\mathrm{d}\rho}{\rho^3\sqrt{\rho(-\rho^2 + \rho\rho_1 - \gamma^2}}\right]\,, \tag{12.21a}$$

where, as before, Y = y/L, L = V_s^2/g, is the scale-height of the atmosphere, and $C(\rho)$ is the cubic,

$$C(\rho) = \rho^3 + 4\pi k^2\rho^2 - \rho(\gamma^2 + 8\pi k^2\rho_1) + 12\pi k^2\gamma^2\,. \tag{12.21b}$$

The above expression can be readily integrated to give Y(ρ) in terms of the incomplete elliptic integrals of the first and second kind, $F(\phi,\kappa)$ and $E(\phi,\kappa)$, respectively. The solution Y(ρ) depends on three constants, namely k, ρ_1 and γ. For any given set of these constants Eq. (12.21) can be integrated to give Y(ρ). In reversing the function Y(ρ) to obtain ρ(Y) some caution is needed however, since for a fixed sign in Eq. (12.21) the function ρ(Y) might not be single-valued if the cubic equation, C(ρ) = 0 has roots ρ_i, i=1,2,3, which happen to be in the interval of the allowed density. In the following we shall examine the allowed solutions of Eq. (12.21) for all possible values of the constants k, ρ_1, and γ.

12.3.3.2 Relations between the Characteristic Speeds

In order to have a qualitative understanding of the solution, in this subsection we investigate the relation of the magnitudes of the three characteristic speeds, the sound speed V_s, the Alfven speed $V_A^{\pm}$ and

the flow speed $V_y^{\pm}$ at the valleys (+) and the summits (−) of the field and streamlines where the density is $\rho^{\pm}$ and the field lines are horizontal.

Conditions at the Valleys

The z-component of the momentum balance equation can be written,

$$\frac{\mathrm{d}}{\mathrm{d}Y}\left[\frac{B_o B_z}{4\pi}\left(1-\frac{V_y^+ V_y}{(V_A^+)^2}\right)\right]=\rho V_s^2\,, \tag{12.22a}$$

or, equivalently,

$$\left[1-\frac{V_y^+ V_y}{(V_A^+)^2}\right]\frac{\mathrm{d}}{\mathrm{d}Y}\frac{B_o B_z}{4\pi}+\frac{B_o B_z}{4\pi}\frac{\mathrm{d}}{\mathrm{d}Y}\left[1-\frac{V_y^+ V_y}{(V_A^+)^2}\right]=\rho V_s^2\,. \tag{12.22b}$$

Noting that at the lowest point of each field line $B_z=0$, $dB_z/dY>0$ and $V_y=V_y^+$, Eq. (12.22b) yields

$$V_y^+ < V_A^+\,. \tag{12.23}$$

This result can be simply understood by writing the equilibrium equations of force balance at the lowest point of the field line

$$\frac{\rho(V_y^+)^2}{r^+}+\rho^+ g=\frac{B_o^2}{4\pi r^+}\,, \tag{12.24a}$$

where r^+ is the radius of curvature at this point. By multiplying by r^+ and dividing by ρ^+ we obtain

$$(V_y^+)^2+r^+ g=(V_A^+)^2\,, \tag{12.24b}$$

from which the inequality (12.23) follows.

On the other hand, consider the y-component of the momentum balance, Eq. (12.18a),

$$\frac{\mathrm{d}\rho}{\mathrm{d}Y}\left[V_s^2-\frac{(\rho^+ V_y^+)^2}{\rho^2}\right]=-\frac{B_z}{4\pi}\frac{dB_z}{dY}\,. \tag{12.25}$$

Taking the derivative with respect to Y :

$$\frac{\mathrm{d}^2\rho}{\mathrm{d}Y^2}\left[V_s^2-\frac{(\rho^+ V_y^+)^2}{\rho^2}\right]+\frac{\mathrm{d}\rho}{\mathrm{d}Y}\frac{\mathrm{d}}{\mathrm{d}Y}\left[V_s^2-\frac{(\rho^+ V_y^+)^2}{\rho^2}\right]$$
$$=-\frac{1}{4\pi}\left[\frac{\mathrm{d}B_z}{\mathrm{d}Y}\right]^2-\frac{B_z}{4\pi}\frac{\mathrm{d}^2 B_z}{\mathrm{d}Y^2}\,, \tag{12.26}$$

and substituting $B_z=0$, $d\rho/dY=0$ at the position of maximum density where the field and stream lines are horizontal we obtain,

$$\frac{\mathrm{d}^2\rho^+}{\mathrm{d}Y^2}\left[V_s^2-(V_y^+)^2\right]=-\frac{1}{4\pi}\left[\frac{\mathrm{d}B_z}{\mathrm{d}Y}\right]^2<0\,. \tag{12.27}$$

Since $\rho^+=\rho_{max}$, we have that $\mathrm{d}^2\rho/\mathrm{d}Y^2<0$ and

$$V_y^+ < V_s\,. \tag{12.28}$$

It follows that the flow at the valleys of the field- and stream-lines where the density is maximum, is subsonic and sub-Alfvenic. This result should not be surprising, since at these localities the equilibrium is dominated by the magnetic tension forces, and therefore we have a Kippenhahn-Schlüter, prominence-like solution with a weak flow along the sagging magnetic field lines.

Conditions at the Summits

Let ρ^-, V_y^-, V_A^- be the density, flow speed and Alfven speed at the summit of some looplike one-dimensional hydromagnetic equilibrium, where the magnetic and stream lines are horizontal. At this point $B_z=0$, $dB_z/dY<0$, and the z-component of the momentum balance equation

$$\left[1-\frac{V_y^- V_y}{(V_A^-)^2}\right]\frac{\mathrm{d}}{\mathrm{d}Y}\frac{B_o B_z}{4\pi}+\frac{B_o B_z}{4\pi}\frac{\mathrm{d}}{\mathrm{d}Y}\left[1-\frac{V_y^- V_y}{(V_A^-)^2}\right]=\rho V_s^2\,, \tag{12.29}$$

yields the following inequality,

$$V_y^- > V_A^-\,. \tag{12.30}$$

As before, this relation can be simply understood by writing the force-balance equation

$$\frac{\rho(V_y^-)^2}{r^-}=\rho^- g+\frac{B_o^2}{4\pi r^-}\,, \tag{12.31}$$

where r^- is the corresponding radius of curvature.

On the other hand, the y-component of momentum balance gives as before [cf. Eq. (12.27)]

$$\frac{\mathrm{d}^2\rho^-}{\mathrm{d}Y^2}\left[V_s^2-V_y^{-2}\right]=-\frac{1}{4\pi}\left[\frac{\mathrm{d}B_z}{\mathrm{d}Y}\right]^2<0\,. \tag{12.32}$$

Since $\rho^-=\rho_{min}$, we have $d^2\rho/dY^2>0$ and the above inequality yields

$$V_y^- > V_s\,. \tag{12.33}$$

The flow at the summits of the field- and stream-lines, where the density is minimum, is supersonic and super-Alfvenic. As before, this result should not come as a surprise, since, for equilibrium, the flow at these localities has to be strong enough that the resulting centrifugal force balances both the plasma weight and the magnetic tension force.

12.3.3.3 1-D MHD Equilibria with Valleys and Summits.

Consider now the question of the existence of such one-dimensional hydromagnetic polytropic equilibria with both valleys and summits. Let us first conveniently choose the constants of the solution according to the analysis of the previous sections. For convenience, redefine the density ρ_1 in terms of the density ρ_o at one of the extrema of the field lines (valleys or summits) where they are horizontal,

$$\rho_1 = \rho_o + \frac{\gamma^2}{\rho_o}. \tag{12.34}$$

Since $B_z^2(\rho) > 0$, it follows that ρ is constrained to lie between the two roots of the quadratic in Eq. (12.20a),

$$\rho^- = \frac{\gamma^2}{\rho_o} < \rho < \rho_o = \rho^+, \qquad \text{case I}, \tag{12.35}$$

if $\gamma < \rho_o$, or ,

$$\rho^- = \rho_o < \rho < \frac{\gamma^2}{\rho_o} = \rho^+, \qquad \text{case II}, \tag{12.36}$$

if $\gamma > \rho_o$. The former case corresponds to setting $\rho = \rho_o$ as the density at the valleys, while the latter corresponds to setting $\rho = \rho_o$ as the density at the summits. This result of a bounded density distribution is a novel feature of 1-D hydromagnetic equilibria in uniform gravitational fields. Apparently it can be regarded as an extension of the known property of 1-D prominence-like, magnetostatic equilibria in uniform gravity where the density is bounded above, $\rho < \rho_o$, but unbounded below [except, of course, by zero]. The additional restriction on the density seems to have a hydrodynamic origin, since it can be seen that the lower bound ρ^- disappears when $V_o \to 0$ (or, equivalently, $k, \gamma \to 0$).

Let $\beta^+ = 8\pi\rho^+ V_s^2/B_o^2$ and $\lambda^+ = V_y^+/V_s$, be the plasma beta and Mach number, respectively, at the valleys of the equilibrium configuration, when $\gamma < \rho_o$, (case I). Eqs. (12.20a), (12.21) for the dimensionless z-component of the magnetic field $H = B_z/B_o$, and the dimensionless density $\mathrm{R} = \rho/\rho^+$ can be written, in terms of the three constants ρ^+, β^+, λ^+,

$$H(R) = \pm\frac{\sqrt{\beta^+[1-R][R-(\lambda^+)^2]}}{R}, \tag{12.37}$$

$$\pm\sqrt{\beta^+}\mathrm{d}Y = \frac{C(R)\mathrm{d}R}{R^3\sqrt{R[1-R][R-(\lambda^+)^2]}}, \tag{12.38a}$$

where

$$C(R) = R^3 + \frac{\beta^+(\lambda^+)^2}{2}R^2 - (\lambda^+)^2[1+\beta^+ + \beta^+(\lambda^+)^2]R + \frac{3\beta^+(\lambda^+)^4}{2}. \tag{12.38b}$$

It is obvious that the density must be in the interval $I = [(\lambda^+)^2, 1]$. In order that we obtain from Eq. (12.38a) a physical solution for $R(Y)$ for a given set of the parameters β^+ and λ^+, we need to examine if the function $Y(R)$ is one-to-one; i.e., we need to check if the cubic $C(R)$ has roots in the interval I. By checking the determinant of $C(R)$ we find that it is negative and therefore $C(R)$ has always three real roots. To check if they fall in the interval I simply note that

$$C(1) = [1-(\lambda^+)^2][1-\beta^+(\lambda^+)^2/2], \tag{12.39a}$$

$$C(\lambda^+) = -\beta^+(\lambda^+)^3[1-(\lambda^+)^2] < 0, \tag{12.39b}$$

$$C((\lambda^+)^2) = (\lambda^+)^4[1-(\lambda^+)^2][\beta^+ - 2]/2, \tag{12.39c}$$

so we have that $2 < \beta^+ < 2/(\lambda^+)^2$, $[V_y^+ < V_A^+ < V_s]$, $C((\lambda^+)^2) > 0$, $C(\lambda^+) < 0$, $C(1) > 0$ and therefore we have two roots in I. At the points R_* where $C(R_*) = 0$, we have that $\mathrm{d}Y/\mathrm{d}R = 0$. However, since the function $R(Y)$ must be single-valued, the solutions must be stopped at this root $R_* < 1$ of $C(R)$ which is closer to 1; and after R_* we may take the mirror symmetric of the curve $Y(R)$ with respect to the line $Y(R_*)$. In doing so we actually take advantage of the $\pm$ signs in expressions (12.37-12.38a).

Altogether, the solution at the valleys is magnetically dominated and the density is of the Kippenhahn-Schlüter type (Section 12.3), while the solution is hydrodynamically dominated at the summits where the density is approximately constant (Section 12.4). The interested reader may find more details on this solution in Tsinganos and Surlantzis (1991).

12.4. Collimated Hydromagnetic Winds

As the second example of constructing symmetric MHD equilibria by using the methodology of section 2, consider the problem of collimated hydromagnetic flows in the atmosphere of a central gravitating

object with mass M, whose gravitational field is g(r),

$$g(r) = -\frac{GM}{r}, \tag{12.40}$$

in spherical coordinates (r, θ, ϕ), with r measured from the central object, and θ from the polar axis. The system is assumed spherically symmetric, i.e., all quantities are independent of the azimuthal angle ϕ. The following choice of the integrals $\Psi(A)$, $\Phi(A)$, and $L(A)$ leads then to physically interesting wind-type solutions,

$$\Psi(A) = \Psi_o + \frac{8\pi\Psi_1}{3\mu}(1+\mu A)^{3/2}, \quad \Phi(A) = \Phi_o + \frac{2\Phi_1}{\mu}(1+\mu A)^{1/2},$$

$$L(A) = -\frac{\Omega_o}{4\pi\Psi_1}\frac{A}{(1+\mu A)^{1/2}}, \tag{12.41a}$$

if the density is expressed in the form,

$$\rho = \frac{\rho_o}{Y(R)R^2}(1+\mu A), \tag{12.41b}$$

where R=r/r_o is the dimensionless distance and ρ_o is the density at the 'base' R=1. If, in addition, the pressure P is separable in the form

$$P(r,\theta) = P_o(r) + P_1(r)A = \frac{B_0^2}{8\pi}Q(r,\theta), \tag{12.41c}$$

and the magnetic flux function A is given by

$$A(r,\theta) = F(r)\sin^2\theta, \tag{12.41d}$$

then equations (12.4) reduce to the following expressions,

$$\mathbf{B} = \left[B_o\frac{f}{R^2}\cos\theta\ ,\ -\frac{B_o}{2R}\frac{\mathrm{d}f}{\mathrm{d}R}\sin\theta\ ,\right.$$
$$\left.\lambda B_o\frac{f}{R}\left(\frac{1-M_A^2Y^*f^*/(Yf)}{1-M_A^2}\right)\sin\theta\right] \tag{12.42a}$$

$$\mathbf{V} = \left[\frac{V_oYf\cos\theta}{[1+\omega f\sin^2\theta]^{1/2}}\ ,\ -\frac{\mathrm{d}f}{\mathrm{d}R}\frac{V_oYR\sin\theta}{2[1+\omega f\sin^2\theta]^{1/2}}\ ,\right.$$
$$\left.\left(\frac{Yf-Y^*f^*}{1-M_A^2}\right)\frac{\lambda V_oR\sin\theta}{[1+\omega f\sin^2\theta]^{1/2}}\right], \tag{12.42b}$$

$$M_A^2 = \left[\frac{V_r}{V_r^A}\right]^2 = \left[\frac{V_\theta}{V_\theta^A}\right]^2 = \frac{YR^2}{\beta^2},\ V_r^A = \frac{B_r}{\sqrt{4\pi\rho}},\ V_\theta^A = \frac{B_\theta}{\sqrt{4\pi\rho}}, \tag{12.42c}$$

and

$$\rho = \frac{\rho_o}{Y(R)R^2}[1+\omega f(R)\sin^2\theta], \quad Q(R,\theta) = Q_o(R) + Q_1(R)\sin^2\theta. \tag{12.42d}$$

With the *ansätz*,

$$A(R,\theta) = F_of(R)\sin^2\theta,$$

the remaining $R-$ and $\theta-$components of the force balance equation (12.1) yield the following equations for $Q_o(R)$, $Q_1(R)$, $Y(R)$ and $f(R)$,

$$Q_1 = \frac{f}{2\beta^2R^2}\frac{\mathrm{d}}{\mathrm{d}R}\left[YR^2\frac{\mathrm{d}f}{\mathrm{d}R}\right] - \frac{Y}{4\beta^2}\left[\frac{\mathrm{d}f}{\mathrm{d}R}\right]^2 - \frac{f}{2R^2}\left[\frac{\mathrm{d}^2f}{\mathrm{d}R^2} - \frac{2f}{R^2}\right]$$
$$-\frac{2\lambda^2}{R^2}\left[\frac{f-[R/R_*]^2f_*}{1-M_A^2}\right]^2 + \frac{\lambda^2}{Y\beta^2}\left[\frac{Yf-Y_*f_*}{1-M_A^2}\right]^2, \tag{12.42e}$$

$$\frac{\mathrm{d}Q_1}{\mathrm{d}R} + \frac{\omega\nu^2f}{\beta^2YR^4} - \frac{2f}{\beta^2R^2}\frac{\mathrm{d}(Yf)}{\mathrm{d}R} + \frac{Yf}{\beta^2R^2}\frac{\mathrm{d}f}{\mathrm{d}R} - \frac{Y}{2\beta^2R}\left[\frac{\mathrm{d}f}{\mathrm{d}R}\right]^2$$
$$+\frac{1}{2R^2}\left[\frac{d^2f}{dR^2} - \frac{2f}{R^2}\right]\frac{\mathrm{d}f}{\mathrm{d}R} - \frac{2\lambda^2}{\beta^2YR}\left[\frac{Yf-Y_*f_*}{1-M_A^2}\right]^2 \tag{12.42f}$$
$$+\frac{\lambda^2}{R^2}\frac{\mathrm{d}}{\mathrm{d}R}\left[\frac{f-[R/R_*]^2f_*}{1-M_A^2}\right]^2 = 0,$$

$$\frac{\mathrm{d}Q_o}{\mathrm{d}R} + \frac{\nu^2}{\beta^2YR^4} + \frac{2f}{\beta^2R^2}\frac{\mathrm{d}(Yf)}{\mathrm{d}R} = 0, \tag{12.42g}$$

where ν and β are the escape speed and the base Alfven speed in units of the radial flow speed at the polar base, respectively,

$$V_{esc} = \sqrt{\frac{2GM}{r_o}}, \quad V_o^A = \frac{B_o}{\sqrt{4\pi\rho_o}}, \quad \nu = \frac{\sqrt{2GM/r_o}}{V_o}, \quad \beta = \frac{V_o^A}{V_o}, \tag{12.42h}$$

and

$$F_o = \frac{B_or_o^2}{2}, \quad \lambda = \frac{\Omega_or_o}{2}, \quad \Phi_1 = -\frac{\Omega_0F_*}{r_*^2}. \tag{12.42i}$$

The three equations (12.42e-g) relate the four quantities $Q_o(R)$,

$Q_1(R)$, $Y(R)$ and $f(R)$. We are free therefore to prescribe one of them. Some such representative cases are presented in the following.

12.4.1 Collimated Hydromagnetic Flows with Prescribed Streamline Shape, f(R)

Consider the simplest case $f(R) = 1$, wherein the field and streamlines are helical. By eliminating Q_1 between Eqs.(12.42e-f) we obtain a single first-order differential equation for $Y(R)$:

$$\frac{\mathrm{d}Y}{\mathrm{d}R} = \frac{F(Y, R; \omega\nu^2; \lambda; \beta; R_*)}{G(Y, R; \omega\nu^2; \lambda; \beta; R_*)}, \tag{12.43a}$$

where

$$F = \frac{\omega\nu^2}{YR^4} - \frac{4\beta^2}{R^5} + \frac{2\lambda^2 Y}{RM_A^2(1-M_A^2)^2}\left[\frac{(2M_A^2-1)}{M_A^2}\frac{R^4}{R_*^4} - (3M_A^2-2)\right], \tag{12.43b}$$

$$G = \frac{2Y}{\beta^2 M_A^2} - \frac{\lambda^2}{(1-M_A^2)^2}\left[\frac{(2M_A^2-1)}{M_A^4}\frac{R^4}{R_*^4} - 1\right]. \tag{12.43c}$$

In the following we examine the fixed points of Eq. (12.43), insofar as they determine the topology and physical behavior of the respective solutions.

12.4.1.1 Critical Points

A topological analysis of Eqs.(12.43) shows that two critical points are present in the solutions, i.e., there are two values of R for which the numerator and denominator of Eq. (12.43a) vanish. These two critical points are introduced by the rotation V_ϕ and the existence of the azimuthal magnetic field B_ϕ.

The first critical point is at $R = R_*$, where $M_A^2 = 1$ ($Y = Y_*$); this is the familiar Alfvenic critical point corresponding to the position where the poloidal flow speed equals the poloidal component of the Alfvén speed. Due to the predominance of terms like $(1 - M_A^2)^{-2}$, for $R \longrightarrow R_*$, the Alfvénic point corresponds mathematically to a high-order singularity, as compared to the ordinary second-order X-type singularity. In fact, by Taylor expanding Eqs. (12.43) at $R \approx R_*$, $Y \approx Y_*$ up to the fourth order, it is found that all slopes are allowed at this Alfvénic point, contrary to the case of the X-type critical point where only two slopes (one positive and one negative) are allowed.

The second critical point is present downstream of R_*, and it is a typical X-type point; its position at R_x can be found through a solution of the system of Eqs. (12.43b-c), $F = G = 0$. Due to the complicated expressions for these two functions F and G, this system can be only solved numerically. A mathematical explanation for the emergence of this second X-type critical point in Eqs. (12.43) has to do with the mathematical nature of the Alfvén high-order singularity, namely that it is of the sink-type which means that there exist infinite solutions that may cross this critical point at R_*. The boundary condition that the flow speed at infinity is finite, in order that the pressure and density at infinity vanish, requires then the existence of the second X-type critical point at R_x such that a single solution meeting that requirement is selected. In the same way this solution must fulfill the requirement that $Y(R = 1) = 1$ with no break in the derivative at R_*; this is achieved by iteratively changing the position of R_* until we have a final critical solution from $Y(R) = 1$ continuously through R_* and R_x, and with high terminal speed.

In Figure 12.1 we plot the topology of a typical solution obtained by fixing ν, ω and λ at some typical solar values, $\nu = 120$, $\omega = 4$, $\lambda = 0.5$ and for $\beta = 100$.

12.4.1.2 Flow Acceleration and Collimation

The acceleration of the wind, depends on collimation through the factor $\omega\nu^2$ and the acceleration and final asymptotic speeds are higher the higher is the collimation, for fixed β and λ. The physical explanation of this result is as follows. At the equator, $\theta = \pi/2$, the plasma weight is balanced by the pressure gradient and the centrifugal force due to rotation. Then, as we move to the pole, $\theta \longrightarrow 0$, the density decreases by the factor ω, while the pressure changes by the factor Q_1 which does not depend explicitly and strongly on ω. Thus, as we move from the equator to the pole, at the same radial distance R, the decrease of the plasma density is larger than the decrease of the pressure. The result is that the acceleration is higher, the higher is the value of ω. This result of the necessity of density increase with the colatitude θ in order to have accelerated outflows, that is shown here with an explicit model, is novel (but see also, Hu and Low 1989, Tsinganos and Trussoni 1990, 1991). Note that in previous hydromagnetic wind models (Low and Tsinganos 1985, Tsinganos and Low 1989) low values for the acceleration and terminal flow speeds were found as a consequence of the assumption of spherically symmetric density, $\omega = 0$.

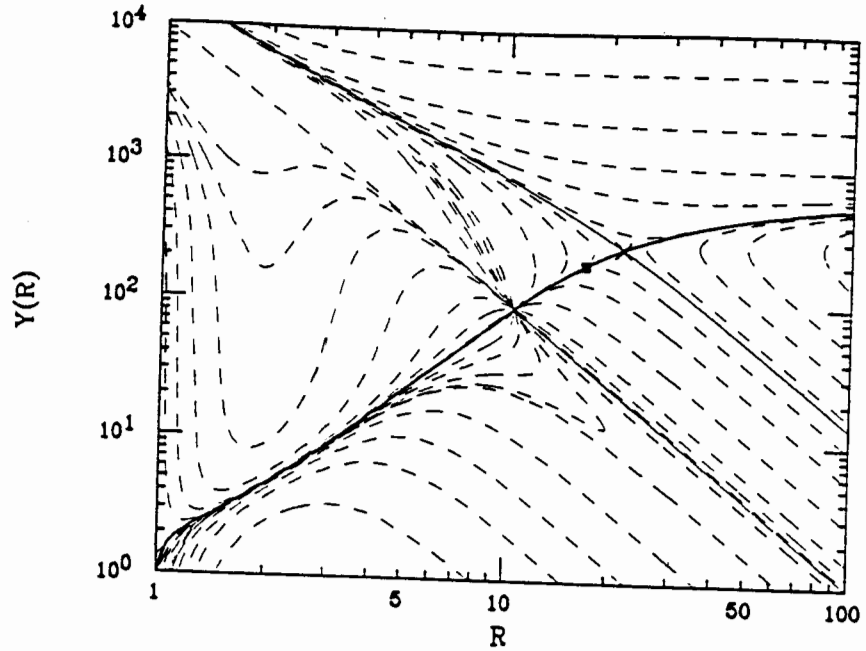


Figure 12.1 The topology of the radial dependence of the radial flow speed $V_R(R,\theta)$ for $\omega = 4$, $\nu = 120$, $\lambda = 0.5$, $\beta = 100$ and $V_o = 5$ kms^{-1}. Note the two critical points, the Alfven point at $R_* = 10.55$ and the X-type point at $R_x = 21.86$. The sonic point is denoted by a square lying between R_* and R_x.

12.4.1.3 Energetics of the Flow

The acceleration of the wind is strictly related to the energy exchange mechanism represented by the function σ. We may write the heating function $\sigma(R,\theta)$, Eq. (12.2b), in the form

$$\sigma(R,\theta) = [\sigma_o(R) + \sigma_1(R)\sin^2\theta]\frac{\cos\theta}{[1+\omega\sin^2\theta]^{3/2}}, \qquad (12.44a)$$

and from Eq. (12.2b) obtain the following expressions for σ_o and σ_1:

$$\sigma_o = \frac{V_o^3 Y}{2r_o(\Gamma-1)}\left[YR^2\frac{\mathrm{d}Q_o}{\mathrm{d}R} + \Gamma Q_o R^2\frac{\mathrm{d}Y}{\mathrm{d}R} + 2\Gamma Q_o YR\right], \qquad (12.44b)$$

$$\sigma_1 = \frac{V_o^3 Y}{2r_o(\Gamma-1)}\left[YR^2\frac{\mathrm{d}Q_1}{\mathrm{d}R} + \Gamma Q_1 R^2\frac{\mathrm{d}Y}{\mathrm{d}R} + 2\Gamma Q_1 YR\right], \qquad (12.44c)$$

where Γ is the ratio of the specific heats ($\Gamma = 5/3$ for a perfect gas). We find that the heating σ is peaked around the position of the critical points while it drops monotonically further away. The polytropic index, Eq. (12.2c), is also representative of the heating distribution along the flow. Regions where $\gamma < 1$ correspond to intense heating, while regions where $\gamma > 1$ correspond to reduced heating. The transition of γ to values below the isothermal value $\gamma = 1$, mark the positions where $\mathrm{d}T/\mathrm{d}R = 0$ and the temperature has an extremum:

$$\gamma = 1 + \frac{\rho}{T}\frac{\mathrm{d}T}{\mathrm{d}\rho}, \qquad \left(\frac{\mathrm{d}\rho}{\mathrm{d}R} < 0\right). \qquad (12.45)$$

In the pure hydrodynamic case, (Tsinganos and Trussoni 1990), we have found that at large R the polytropic index γ approaches asymptotically the Parker value 3/2, since, the density varies as R^{-2} and the pressure as R^{-3}. In this case however, the pressure drops as R^{-2} similarly to the density, with the result that the heating creates an isothermal atmosphere. Thus, in magnetically dominated outflows, an extended nonmonotonic heating is required which results in higher terminal speeds. This is exactly the opposite from what we found in the corresponding hydrodynamically dominated outflows, where a monotonic heating peaked at the base resulted in relatively low terminal speeds.

12.4.2 Collimated Helicoidal Flows with an Azimuthal Magnetic Field, β=0

A simple subcase of the previous helicoidal wind pattern is obtained if β=0, i.e., when there is only an azimuthal component of the magnetic field present, while there is no poloidal field,

$$V_R(R,\theta) = V_o Y(R)\frac{\cos\theta}{[1+\omega\sin^2\theta]^{1/2}},$$
$$V_\phi(R,\theta) = \frac{\lambda V_o}{R}\frac{\sin\theta}{[1+\omega\sin^2\theta]^{1/2}}, \qquad (12.46a)$$

$$B_\phi(R,\theta) = \frac{B_1}{YR}\sin\theta, \qquad \rho(R,\theta) = \frac{\rho_o}{Y(R)R^2}[1+\omega\sin^2\theta], \qquad (12.46b)$$

The r- and θ-components of the momentum balance equation simplify considerably in this case,

$$Q_1 = \frac{\lambda^2}{YR^4} - \frac{2\Lambda^2}{Y^2R^2}, \qquad (12.47a)$$

$$\frac{dQ_1}{dR} = -\frac{\omega\nu^2}{YR^4} + \frac{2}{R^2}\frac{dY}{dR} + \frac{2\lambda^2}{YR^5} + \frac{2\Lambda^2}{Y^3R^2}\frac{dY}{dR}, \qquad (12.47b)$$

$$\frac{dQ_o}{dR} = -\frac{\nu^2}{YR^4} - \frac{2}{R^2}\frac{dY}{dR}, \qquad (12.48)$$

where $\Lambda(= B_1/\sqrt{4\pi\rho_o}V_o)$ is the ratio between the Alfven speed associated with the amplitude of the azimuthal magnetic field B_1 and the radial flow speed at the polar base. By eliminating Q_1 between Eqs.(12.47) we obtain a single first order differential equation for $Y(R)$:

$$\frac{dY}{dR} = \frac{Y}{R}\frac{\omega\nu^2R - 6\lambda^2 + 4\Lambda^2R^2/Y}{2Y^2R^2 + \lambda^2 - 2\Lambda^2R^2/Y}, \qquad (12.49)$$

It is evident that for some range of the parameters this differential equation for Y(R) has a critical point at R = R_x where both the numerator and the denominator vanish simultaneously. With the usual analysis around this critical point, $Y = Y_x(1+\delta)$, $R = R_x(1+\epsilon)$, a second-order equation for the slope $s = \delta/\epsilon$ at R_x is found, with the two roots $s_{1,2}$. For each set of the parameters ω, λ and Λ the location R_x of the critical point is found together with the value of $Y(R = R_x) = Y_x$. For example, when $\omega = 0$, the two slopes $s_{1,2}$ of Y(R) at $R = R_x$ are,

$$s_{1,2} = \frac{-2 \pm \sqrt{88}}{7}, \qquad (12.50)$$

one positive and one negative, resulting in the familiar X-type critical point. At this critical point R_x and Y_x are,

$$R_x = \left(\frac{3\lambda}{2}\right)^{1/3}\left(\frac{\lambda}{\Lambda}\right)^{2/3}, \quad Y_x = \left(\frac{2\lambda^2}{3}\right)^{1/3}\left(\frac{\Lambda}{\lambda}\right)^{2/3}, \quad Y_xR_x = \lambda. \qquad (12.51a)$$

The critical point at (R_x, Y_x) is the point where the total flow speed, V, equals the fast mode wave speed, V_f,

$$V^2 = V_R^2 + V_\phi^2 = V_f^2 = V_s^2 + V_A^2, \quad V_s^2 = \gamma\frac{P}{\rho} = \frac{\gamma QYR^2}{2}, \quad \gamma = \frac{3}{2}. \qquad (12.51b)$$

Note that the boundary condition $Y(1) = 1$ combined with the requirement that the solution crosses the X-type critical point at (R_x, Y_x) leads to an eigenvalue-type condition for the determination of one of the constants ω, λ, and Λ. More details on this solution may be found in Tsinganos and Trussoni (1991).

12.4.3 Determination of Streamline Shape f(R) in MHD Collimated Flows

Sometimes it is of interest to determine the streamline shape, $f(R)$. This can be achieved by closing the system of equations (12.42e-g) with an extra relationship between $Q_o(R)$, $Q_1(R)$, $Y(R)$ and $f(R)$. To illustrate the possibilities we shall outline a case wherein such a relationship is imposed between the two pressure components P_o and P_1,

$$P_1(R) = \kappa P_o(R)F(R), \qquad (12.52a)$$

where κ is a constant (Tsinganos and Low 1989). This is equivalent to saying that the temperature has the same basic radial profile along each line of force along which A(r, θ) is a constant,

$$T = \frac{m_p}{2k}\frac{P_o}{\rho}(1 + \kappa A). \qquad (12.52b)$$

Eqs. (12.42e-g) may then be solved for any given set of the six parameters ν, ω, β, λ, κ and s = df(R=1)/dR (Tsinganos and Sauty 1991). Several special cases may be examined separately in order to see the effects of each parameter in the acceleration of the wind.

12.4.3.1 Rotating and Collimated Hydrodynamic Flows, $\beta = 0$

Setting $\beta = 0$, a rotating (if $\lambda \neq 0$) collimated (if $\omega \neq 0$) hydrodynamic flow is obtained with a nonspherically symmetric pressure (if $\kappa \neq 0$). Note that s=0 corresponds to an initially radial outflow, s < 0 corresponds to an initial positive flaring of the streamlines, while s > 0 corresponds to an initial negative flaring, i.e., convergence of the streamlines towards the pole. In other words, the cases s < 0 have a coronal hole-type geometry while the s > 0 cases have a jet-type geometry. It is remarkable that an analytical expression for the streamline shape function f(R) can be found if $\kappa = 0$. If κ=ω=0, and s > 0 the flow is accelerated while if s<0 the flow is decelerated. In Figures 12.2a,b the shape of the streamlines, from a selfconsistent solution of Eqs. (12.42e-g)-(12.52), in a flow that starts with positive/negative flaring at the base is shown. Note the attraction of the streamlines towards the flow axis, the subsequent conical shape of streamlines and the associated high terminal speed, reminiscent of jet-type flows that is obtained in the example of Figure 12.2b.

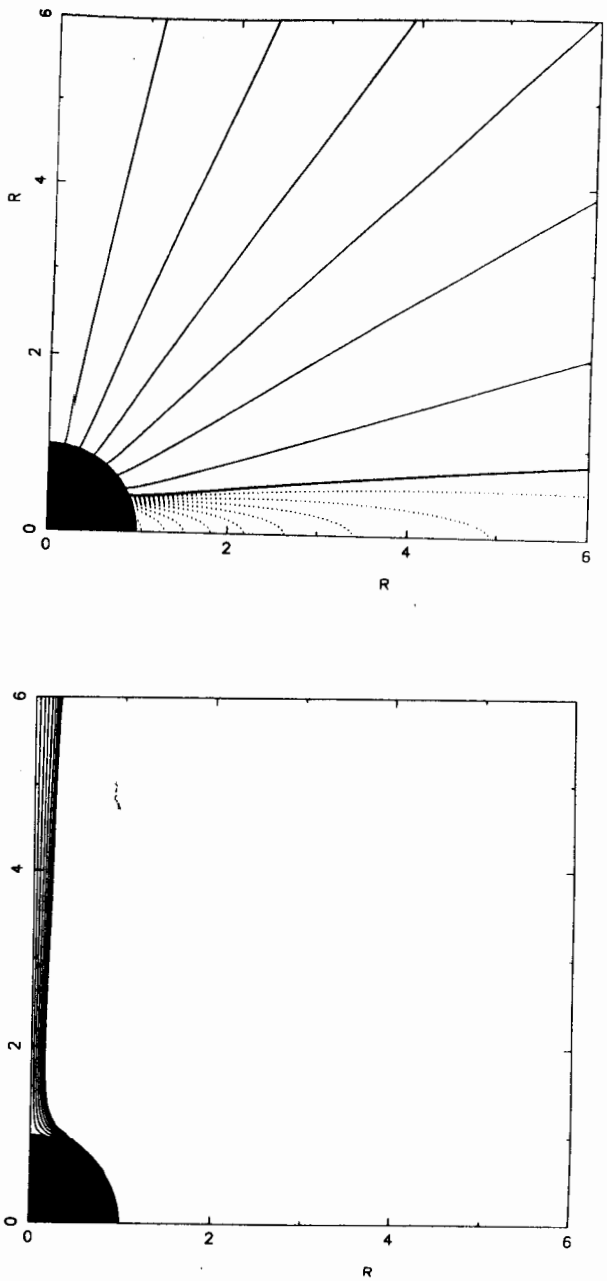


Figure 12.2 Streamline pattern for an initially flaring hydrodynamic outflow, s = -0.830, and $\omega = 2.14$, $\nu = 5$, $\lambda = 0.001$, $\beta = \kappa = 0$ in (a) and for a flow with a negative initial flaring, s = 49, $\omega = 50$ and same ν, λ, β and κ, in (b). The thick streamline separates the open (solid) streamlines around the rotation axis from the closed (dotted) streamlines around the equator.

12.4.3.2 Rotating and Collimated Hydromagnetic Flows, $\beta \neq 0$

The most general case, $\beta \neq 0$, $\omega \neq 0$, $\lambda \neq 0$, $\kappa \neq 0$ and s $\neq 0$, is of considerable interest. The effects of the poloidal and azimuthal magnetic fields, collimation, rotation, nonspherically symmetric pressure and temperature distribution and finally deviations of the field line geometry from radiality at the base may be studied separately in this case.

References

de Ville, A., Priest, E.: 1990, *Solar Physics*, submitted.
Hu, Y.Q., Low, B.C.: 1989, *Astrophys. J.* **342**, 1049
Kippenhahn, R., Schlüter, A: 1957, *Zs. f. Ap.*, **43**, 36
Low, B.C.: 1975 , *Astrophys. J.* **197**, 251
Low, B.C., Hundhausen, A.J., Zweibel, E.G.: 1983, *Phys. Fluids*, **26**, 2731.
Low, B.C., Tsinganos, K.: 1986, *Astrophys. J.*, **302**, 163
Parker, E.N.: 1963, *Interplanetary Dynamical Processes*, Interscience, New York
Parker, E.N.: 1979, *Cosmical Magnetic Fields*, Clarendon Press, Oxford
Priest, E.: 1984, *Solar Magnetohydrodynamics*, D. Reidel, Dordrecht
Tsinganos, K.: 1981, *Astrophys. J.* **245**, 764
Tsinganos, K.: 1982, *Astrophys. J.* **252**, 775
Tsinganos, K., Distler, J., Rosner, R.: 1984, *Astrophys. J.* **278**, 409
Tsinganos, K., Low, B.C.: 1989, *Astrophys. J.* **342**, 1028
Tsinganos, K., Sauty, C.: 1991, *Astron. Astrophys.* submitted.
Tsinganos, K., Surlantzis, G.: 1991, *Solar Phys.*, submitted.
Tsinganos, K., Trussoni, E.: 1990, *Astron. Astrophys.* **231**, 270.
Tsinganos, K. Trussoni, E.: 1991, *Astron. Astrophys.*, in press.

Lothian, R M & Hood, A W (1990) in preparation.
Low, B C (1977) *Astrophys. J.* **217**, 988.
Low, B C (1988) *Astrophys. J.* **330**, 992.
Melville, J P, Hood, A W & Priest, E R (1983) *Solar Phys.* **87**, 301.
Melville, J P, Hood, A W & Priest, E R (1984) *Solar Phys.* **92**, 15.
Melville, J P, Hood, A W & Priest, E R (1986) *Solar Phys.* **105**, 291.
Melville, J P, Hood, A W & Priest, E R (1987) *Geophys. Astrophys. Fluid Dynamics* **39**, 83.
Mikic, Z, Schnack, D & Van Hoven, G (1990) *Astrophys. J.* submitted.
Raadu, M (1972) *Solar Phys.* **22**, 425.
Robertson, J A, Hood, A W & Lothian, R M (1990) in preparation.
Schindler, K, Birn, J & Janike, L (1983) *Solar Phys.* **87**, 103.
Steinolfson, R & Tajima, T (1987) *Astrophys. J* **322**, 503.
Van der Linden, R, Goossens, M & Kerner, W (1990) *Computer Phys. Comm.* **59**, 61.
Velli, M, Einaudi, G & Hood, A W (1990) *Astrophys. J.* **350**, 428.
Zweibel, E (1981) *Astrophys. J.* **249**, 731.
Zweibel, E & Boozer, A (1985) *Astrophys. J.* **295**, 642.
Zwingmann, W (1987) *Solar Phys.* **111**, 309.

15

MHD Turbulence in the Solar Wind

A. MANGENEY, R. GRAPPIN, M. VELLI*

Observatoire de Paris-Meudon F-92190 Meudon France

*European Space Agency Postdoctoral Research Fellow

15.1 Introduction

The Sun has to adjust to an environment, the interstellar medium, which is at negligible pressure. Parker (1963) showed that it is impossible to maintain hydrostatic equilibrium between the hot solar corona and interstellar space. This nonequilibrium drives the solar wind, which is a supersonic, low-density flow. At one astronomical unit, the velocity U is between 300 and 800 km/s, the proton density n about 10 cm^{-3}, the electronic temperature $T_e \approx 1.5 \cdot 10^5$ °K, the proton temperature $T_p \approx 10^5$ °K, and the magnetic field $B \approx 10^{-4}$ G. The solar surface is characterized by inhomogeneities distributed over a wide range of scale lengths. As a result, the wind develops a complex structure. Schematically, however, it may be described as being made up of two high-velocity, low-density, hot flows originating from the polar coronal holes, and separated by a relatively thin sheet of dense, cold and low-velocity wind, located approximately in the vicinity of the ecliptic plane. It is permeated by a large-scale magnetic field, of the same polarity in each region of fast wind as that of the coronal hole from which it emerges. The magnetic neutral sheet which separates these two polarities is embedded in the region of dense cold plasma which separates the two streams. When projected towards the surface of the Sun, it appears that this neutral sheet and the associated low-speed wind originate from the regions of complex magnetic field structures, made of closed field lines and "helmet streamers" which lie above the neutral line separating regions of opposite polarity at the photospheric level. The warping of the neutral sheet (the so-called ballerina skirt) around the equator induces in the ecliptic plane sectors of opposite magnetic polarities which are seen by a spacecraft. Due to the rotation of the solar surface and to

the frozen-in property of the magnetic field, the boundaries between the sectors take the form of a spiral.

This picture of the solar wind is valid near solar minimum. For other periods, the situation is more complex; for example, the geometry of the magnetic neutral line can become highly distorted. Purely thermal driving of the Solar Wind can only explain the slowest flows, with 300 km/s $\leq$ U $\leq$ 400 km/s (Leer, Holzer, Fla, 1982), i.e., it explains the "non typical" wind in the sheet and not the bulk of the flow. Furthermore, the proton temperature usually decreases much more slowly with heliocentric distance r than predicted by adiabatic expansion of the plasma, even if the conductive energy flux is negligible at 1 AU. Hence, non-thermal addition of momentum and energy is needed beyond the critical points, in the supersonic region. The most obvious way to transport energy and momentum at large distances is by waves (Alazraki and Couturier, 1971; Jacques, 1977). Among the various modes which may be invoked, Alfvén waves play a special role, since at least in simple and homogeneous configurations, they are the least affected by damping (see Barnes 1979). Fluctuations resembling Alfvén waves have indeed been observed in the solar wind and the dynamics of their interaction and the relation with MHD turbulence will be the main subject of this review.

In addition to the heating and acceleration problem, the great interest of solar wind MHD turbulence is that it is the only example of a high (kinetic and magnetic) Reynolds number turbulent flow which can be measured in situ. This may appear paradoxical, since the collisional (Coulomb) mean free path λ is comparable to the size L of the system: $\lambda \approx L \approx 1$ AU, and the turbulent velocities u are comparable to the thermal proton speed c_s so that the corresponding Reynolds number is about unity: $Re \approx uL/c_s\lambda \approx 1$ (implying in this case that a fluid treatment is not obviously applicable). In fact, wave-particle interactions lead to a much smaller effective mean free path λ, of the order of the Larmor radius. This justifies a fluid description in which the effective viscosity is small, and the effective Reynolds number large, in agreement with the high number of degrees of freedom observed (see also Montgomery (1983), and the recent review by Marsch (1990a) for the dissipation mechanisms on the solar wind). Hence, the astrophysical problem of heating the solar wind is intimately related to the more general physical problem of nonlinear interactions and MHD turbulence, since the energy has to be efficiently transferred from large to small (proton gyroradius) scale.

In this review we shall concentrate on the inner heliosphere (R $\leq$ 1 AU) because it is here that the heating and acceleration problems are the most severe and also the departure from standard stationary turbulence more intriguing. For a more complete picture of the subject, see for instance the papers of Hollweg (1986), Roberts et al (1989), and the reviews by Barnes (1983), Isenberg (1989), Marsch (1990b).

The chapter is organized as follows: in Section 15.2 the basic characteristics of the wind necessary to understand the in situ measurements are described. In Section 15.3, we discuss how and to what extent the data in a medium frequency range support the idea of evolution of MHD turbulence. In Section 15.4 we discuss the specificity of Alfvénic turbulence versus "standard" MHD turbulence in the context of the solar wind.

15.2 The solar wind as a wind tunnel flow

15.2.1 The role of solar magnetic field structures

The interplanetary medium can be considered to some extent as being statistically stationary in time (see e.g., Matthaeus and Goldstein 1982; Matthaeus et al. 1986), but not in space, because of the spherical expansion of the wind: as a result, there is no reason to expect the turbulence to have the same characteristics near and far from the Sun.

The spatial distribution of magnetic fields on the Sun, with a variable mixture of open and closed magnetic structures is another cause of inhomogeneity: the proportion of open field lines is larger in the polar coronal holes than near the equatorial region (the equatorial zone expands during solar maximum activity). These magnetic structures play two roles. First, because they are continuously shaken by underlying photospheric or sub-photospheric motions, they are a source of MHD waves. Second, because the closed structures inhibit the birth of the wind, the alternance of open and closed configurations provides a kind of grid for the wind near the Sun. The MHD waves will be found at larger distances in the wind if they are able to propagate through the highly inhomogeneous solar atmosphere: the exact location of the wave source and the extent of the filtering by the atmosphere are still controversial (for differing points of view see Hollweg 1978, Leroy 1981) and one must admit

that the initial conditions at the entry of the wind tunnel, and especially the energy spectrum of the fluctuations, are not known.

The magnetic field lines eventually close within the subsonic part of the flow, between the photosphere and 2 solar radii. Immediately above, in the wake of the closed structures, one should find slower wind velocities than above open structures. The existence of the resulting collection of streams with various velocities can be inferred from observations of the interplanetary scintillation of cosmic radiosources (Scott et al. 1983). These velocity fluctuations are found to culminate at about 10-20 solar radii, where the wind also appears to approach a substantial fraction of its terminal velocity.

Further from the Sun, these separated streams merge, thus reducing the tangential gradients in radial velocities; the same observations by Scott et al. show indeed that the variance of the radial velocity decreases beyond the 10-20 solar radii, while the average velocity keeps increasing slowly. After the merging of these middle scale structures, there still remain the two main high speed polar streams already described above, and a slow wind in the wake of the densest part of the magnetic "grid". Since the solar magnetic neutral line is warped around the equator, one finds an alternance of fast and slow winds in the ecliptic plane. The rotation of the solar source provides a supplementary means to develop the turbulence further downwards from the grid, because fast streams will interact with the slow streams and create shocks in the process. The interaction at such large scales (the progressive merging of slow and fast winds via shock waves) have been observed at very large heliocentric distances in the Voyager data (Burlaga 1987). The main difference between the inner ($r \leq 1$ AU) and the outer heliosphere is thus the following: in the latter region, the largest scales i.e. the sector and stream structures, will have injected a substantial part of their energy in the smaller scales, in the form of shocks and interaction regions which fill a large part of the space, at least in the explored part, near the ecliptic.

Finally, the geometry of the solar wind tunnel is a diverging one: the resulting cooling of the medium, and the associated energy transfer from the waves to the bulk flow will have significant effects on the evolution of the turbulence. In particular, the solar wind velocity fluctuations are supersonic, not only in the close vicinity of the source, but in the whole heliosphere. This makes it very different from the usual numerical or experimental situation, in which the

turbulent Mach numbers decrease very rapidly with time or distance from the source (see Passot and Pouquet, 1987).

15.2.2 A synopsis of in situ measurements

What information is really contained in the power spectra depends on the frequency range. When the frequency is larger than about 1 day^{-1} (corresponding to the travel time from the Sun to a spacecraft located at about 0.3 AU), one may use Taylor's hypothesis to translate frequency f in the satellite frame of reference into absolute length scales or wave vector k by $f = kU/(2\pi)$, since the wind speed U is much larger (10 times at 1 AU) than both the Alfvén (V_a) and thermal (c_s) speeds most of the time. This is true except for the range between 10 and 1000 Hz where the fluctuations take the form of whistler waves. Apart from this range, the fluctuations are convected essentially radially, so that the basic information is that of the radial spectrum of the fluctuations at a given distance from the Sun (this is not strictly correct, since the source rotates: the tangential structure is also recorded in an indirect way in the data, and hence the measured spectra are not purely radial).

Below one day^{-1}, the satellite records are a mixture of temporal variations and of azimuthal variations, and also of pure time-variations of the solar initial conditions. If the spacecraft (as in the case of Helios and Voyager) is not on an earth-bound trajectory, one obtains information on the variation of the spectra with heliocentric distance. The solar wind has been explored in a wide range of heliocentric distances from 0.3 to 40 AU, but only in the vicinity of the ecliptic plane and along the (special) trajectory of a spacecraft. These conditions are not representative of the whole solar cavity: direct observations of the solar wind outside the ecliptic plane will be made possible by the Ulysses spacecraft which should give us a much better understanding of typical solar wind conditions. Figure 15.1 presents a schematic view of the power spectrum of the different types of fluctuations as measured at 1 AU. One of the most striking features of solar wind is the very large number of degrees of freedom which are excited: fluctuations are observed on all time scales from 10^{-6} Hz up to 10^4 Hz. The figure shows the average specific energy spectrum of the fluctuations of the magnetic field δB, the plasma velocity u, the proton density ρ and the electric field E observed in the solar wind plasma at about 1 AU. (See also Denskat and Neubauer

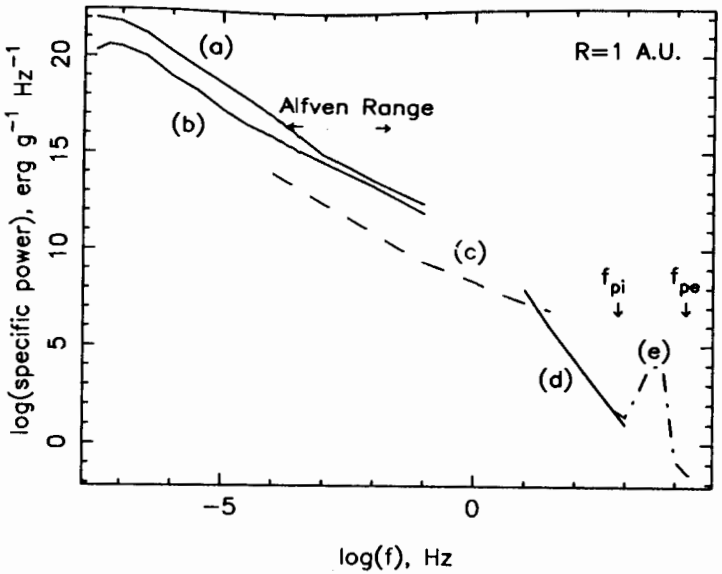


Figure 15.1. An outline of the spectrum of excitations found in the solar wind. Energy per unit mass per Hz is plotted vs. frequency in the satellite frame: (*a*), $\frac{u^2}{2}$; (*b*), $\frac{b^2}{2}$; (*c*), $\left(\frac{c_s \delta n}{n}\right)^2$; (*d*), $\frac{b^2}{2}$ (whistler); (e), electrostatic energy density

(1983), figure 7, for a power spectrum of the magnetic fluctuations in the frequency range 10^{-5} to 10^{2} Hz).

At frequencies higher than 1 Hz (the typical proton gyrofrequency) wave-particle interactions dominate the spectral transfer and effectively dissipate the energy. The associated spatial scale is the proton gyroradius $l_g \approx 100$ km. Between 1 Hz and the proton plasma frequency f_{pi} ($\leq$ 1 kHz), dispersive whistler waves are observed with a power spectrum in $f^{-3.4}$. At still higher frequencies, the modes become essentially electrostatic and have been interpreted as ion acoustic waves.

15.3 Is there a turbulent cascade in the solar wind?

15.3.1 Evidence for an inertial domain

The shape of the velocity and magnetic field spectra shown in Figure 15.1 are strongly suggestive of a nonlinear cascade, where one expects to find energy on all possible wavevectors, although not uniformly distributed: a specific property of turbulence is the scale invariance of the energy spectrum, which manifests itself in the form of a power-law spectrum. This viewpoint was first adopted by Coleman (1968).

The large scales show a dominance of kinetic energy over magnetic energy ($f \leq 10^{-4}$ Hz) while at smaller scales both kinetic and magnetic energies are approximately of the same order of magnitude. This is true up to the highest frequencies at which the solar wind velocity has been measured, which is about $(1mn)^{-1}$. In this last range, the spectral index is often around 1.5-1.7: these values are compatible with the spectral slopes that one expects on the basis of either fluid or MHD turbulence (as will be shown below). Moreover, density fluctuations $\delta\rho$ are usually weak, $\delta\rho/\rho \ll u/c_s$ (where c_s is the proton thermal speed and u is the rms velocity fluctuation): this means that the amplitude of compressible waves is small. It is this range, between 10^{-4} Hz and 10^{-2} Hz, which we shall call the Alfvén range. On the other hand, Belcher and Davis (1971) observed that, during a substantial portion of the time (the so-called "Alfvénic periods"), the velocity $\mathbf{u}$ and the magnetic field fluctuations $\mathbf{b}$ not only are of the same magnitude, but are almost completely correlated: either $\mathbf{z}^+ = \mathbf{u} + \mathbf{b}$ or $\mathbf{z}^- = \mathbf{u} - \mathbf{b}$ (where $b = \delta B/\sqrt{4\pi\rho}$ is the magnetic field fluctuation in velocity units), is often very small, depending on the polarity of the average field $\mathbf{B}^\circ$, so that the waves propagating away from the Sun appear to dominate (Figure 15.2). This suggests that the fluctuations consist of linearly propagating waves, a possibility confirmed (Dobrowolny, Mangeney, Veltri 1980a) by inspection of the incompressible MHD equations in a uniform field in static equilibrium, which may be written formally as

$$\frac{\partial \mathbf{z}^\pm}{\partial t} \mp \mathbf{V}_a \cdot \nabla \mathbf{z}^\pm = -\frac{1}{\rho}\nabla p^T - (\mathbf{z}^\mp \cdot \nabla \mathbf{z}^\pm) \qquad (15.1a)$$

where $\mathbf{V}_a = \mathbf{B}^\circ/(4\pi\rho)$ is the Alfvén velocity and p^T is the fluctuation in the total (kinetic and magnetic) pressure; introducing the Fourier amplitudes of $\mathbf{z}^\pm$, $\mathbf{z}^\pm_\mathbf{k}$, the Fourier transform of equ. (15.1a) may be written as

$$\left(\frac{dz^\pm_\mathbf{k}}{dt} \mp i\mathbf{k}\cdot\mathbf{V}_a z^\pm_\mathbf{k}\right) = -\frac{1}{2} i\mathbf{P}(\mathbf{k}) \int_{\mathbf{p}+\mathbf{q}=\mathbf{k}} \mathbf{z}^\mp_\mathbf{p}\, \mathbf{z}^\pm_\mathbf{q}\, d^3\mathbf{q} \qquad (15.1b)$$

where the tensor $\mathbf{P}(\mathbf{k})$ has components

$$\mathbf{P}(\mathbf{k})_{ilm} = k_m\left(\delta_{il} - \frac{k_i k_l}{k^2}\right) + k_l\left(\delta_{im} - \frac{k_i k_m}{k^2}\right) + k_l\delta_{im} - k_m\delta_{il}.$$

It is clear from equ.(15.1b) the nonlinear interactions vanish when either z^+ or z^- is zero. The existence in the solar wind of a well established, scale-invariant spectrum made up of non-interacting waves

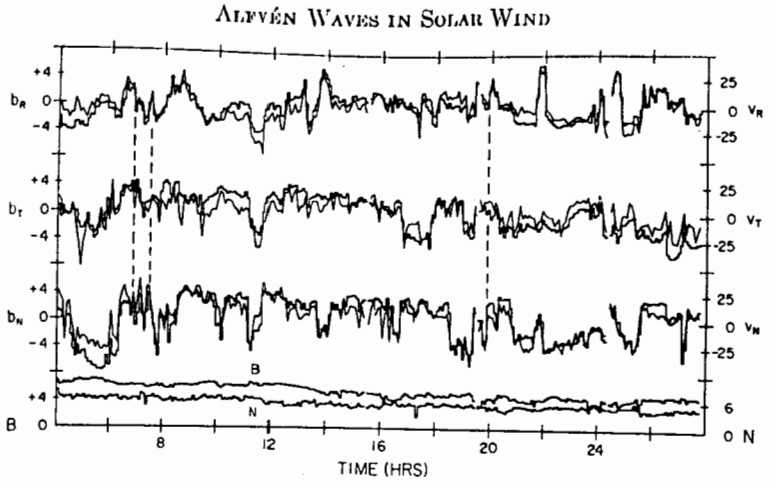


Figure 15.2. A celebrated picture by Belcher and Davis (1971) showing a one-day period with particularly well correlated velocity and magnetic fluctuations.

is problematical, as remarked by Dobrowolny et al. (1980a): given a solar source of outgoing waves, the observed spectrum should reflect its properties, as well as the filtering of the intervening medium, and hence gaps in the spectrum or peaks at some typical generation frequencies (or harmonics) should appear, which are not observed in the data (Fig. 15.1 and 15.2). However the standard situation is that of a mixture of both types of Alfvén waves, with moderate dominance of outward propagating waves (about a factor 4 in the average energy ratio, see figure 15.3), so that the nonlinear couplings are not vanishingly small. In this Section we shall concentrate on this quasi-symmetric regime ($z^+ \approx z^-$), which one may attempt indeed to describe by 'standard' MHD theory, as first proposed by Coleman (1968), and come back in the next Section to the specific properties of the asymmetric ($z^+ \gg z^-$) state, which is typical of Alfvénic periods.

First, a discussion of the time scales is in order. Just as in any wind tunnel, one cannot expect to observe fully developed turbulence too close to tunnel entry; some time must be allowed for the non-linear effects to develop significantly in the plasma convected away from the Sun. This time is the turnover-time τ_{nl}, which for an eddy of size l depends on the velocity fluctuation u within the eddy: $\tau_{nl} \approx l/\mathrm{u} = (\mathrm{U}/\mathrm{u})\mathrm{T}$, if $\mathrm{T} = l/\mathrm{U}$ is the period measured in the spacecraft frame. At a given heliocentric distance r, the transport time by the

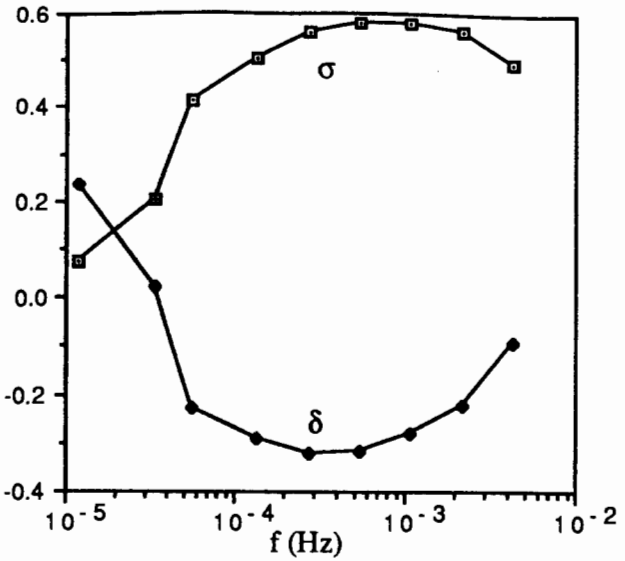


Figure 15.3. In the "Alfvén spectral range" the outward/inward ratio z^+/z^- is significantly larger than unity (parameter $\sigma = (z^{+2} - z^{-2})/(z^{+2} + z^{-2})$) and magnetic energy dominates over kinetic energy (parameter $\delta = (u^2 - b^2)/(u^2 + b^2)$). This figure shows averages made at solar minimum (during the first three months of the Helios 1 mission), at heliocentric distances between 0.3 and 1 AU (from figure 2 in Grappin et al, 1989).

average flow is $\tau_{tr} = r/\mathrm{U}$. A third time scale, $\tau_{ad} \approx 1/(\nabla\cdot\mathbf{U}) = -(1/\rho)\mathrm{D}\rho/\mathrm{Dt}$, describes the rate of change of the plasma specific volume $1/\rho$ associated with the geometry of the expansion, and does not depend on the spatial scale of the eddies. The adiabatic and transport times are of the same order of magnitude in the supersonic region of the wind (they are identical for a spherical expansion with constant speed). For example at r =0.3 AU, the distance of closest approach of the Sun by the Helios spacecrafts, $\tau_{tr} = r/\mathrm{U}$ =35h; an inspection of figure 1 shows that in the Alfvén range, with U about 600 km/s, u/U ≈ 0.05, and thus $\tau_{nl} < \tau_{tr} \approx \tau_{ad}$. If we follow a plasma parcel which is convected with the wind, nonlinear effects will affect strongly only the part of the spectrum for which the average turnover time is smaller than the transport time. Therefore the width of the inertial range, if it does exist, must depend on the radial distance. Except if the spectrum is very steep, the nonlinear time decreases with the scale; hence there will always be a critical scale $\mathrm{L}(r) \approx (\mathrm{u}/\mathrm{U})r$ below which nonlinear effects dominate (Tu et al. 1984). As heliocentric distance r increases, the adiabatic time also increases, and so does $\mathrm{L}(r)$, as long as u does not decrease as fast

as r^{-1}, i.e., as long as the turbulent specific energy does not decrease as fast as r^{-2}. The nonlinear interactions are thus free to redistribute the energy among the degrees of freedom available between the scale L(r) and a dissipation scale $l_d \approx l_g$.

15.3.2 Phenomenology of homogeneous, isotropic and incompressible turbulence

A fully developed turbulent state is expected when several orders of magnitude separate the two scales (L(r) $\gg$ l_d). According to the Kolmogorov (1941) theory, two properties characterize such a state. First, the energy dissipation rate is independent from the viscosity of the fluid, i.e., it reaches a finite value in the limit of zero viscosity. Since the nonlinear interactions respect the conservation of energy, the dissipation rate must be given by the energy injection rate $\Pi^\circ = \epsilon$, which either comes from an external source, or from the largest, energy-containing eddies (here at scale L(r)). Second, the energy is not transfered directly from the largest scale down to the dissipation scale, but instead is transfered via successive interactions between smaller and smaller (but at each step comparable) wavenumbers, (i.e., the dominant contribution to the rhs of eq.(15.1b), comes from interactions with $|\mathbf{p}| \approx |\mathbf{q}| \approx |\mathbf{k}|$) whence the name "energy cascade". During this cascade, eddies of a given size l break into smaller eddies, but are regenerated by the larger eddies, and so on; since all scales are in energetic equilibrium, the energy dissipation rate of eddies of size l, $\Pi(l)$, is independent of the scale l, i.e., it is equal to ϵ : $\Pi(l) = \Pi^\circ = \epsilon$. Let $\tau^*(l)$ be the characteristic time for an eddy of size l to transfer its energy E(l), so that

$$\Pi(l) \approx \mathrm{E}(l)/\tau^*(l). \qquad (15.2)$$

Since the energy transfer in an ordinary fluid results essentially from the self-distortion of the eddy, the transfer time is simply equal to the turnover-time τ_{nl}. Assuming that the interactions are local in wavenumber space, it may be written as $\tau_{nl} \approx l/\mathrm{u}(l)$, where u($l$) is the characteristic velocity dispersion within the eddy. Using $\mathrm{E}(l) \approx \mathrm{u}^2(l)$, we finally obtain that $\mathrm{E}(l) \approx (\epsilon)^{2/3}$, or since (with $\mathrm{k} \approx 1/l$) $\mathrm{E}(l) \approx \int^{\mathrm{k}} \mathrm{E}_{\mathrm{k}'}\mathrm{d}\mathrm{k}' \approx \mathrm{k}\mathrm{E}_\mathrm{k}$:

$$\Pi_\mathrm{k} \approx \mathrm{k}^{5/2}\mathrm{E}_\mathrm{k}^{3/2} = \epsilon \qquad (15.3)$$

and the scale-invariance of the flux leads then to $\mathrm{E}_\mathrm{k} \approx \epsilon^{2/3}\mathrm{k}^{-5/3}$, which is the Kolmogorov spectrum.

Now, consider the case of a conducting fluid embedded within a uniform magnetic field, and assume that there are incompressible fluctuations, in the form of Alfvén waves. The basis of Iroshnikov's (1963) and Kraichnan's (1965) (henceforth IK) theory lies in the fact that self-distortion, in a large-scale magnetic field, is replaced by weaker interactions between propagating Alfvén waves. Propagation introduces an additional timescale, the Alfvén time $\tau_\mathrm{A} \approx l/\mathrm{V_a}$, so that the effective energy transfer time τ^* is no longer equal to the eddy-turnover-time $\tau_{nl} \approx l/\mathrm{u} \approx l/\mathrm{b}$. Since to lowest order the nonlinear terms couple linear solutions (i.e. Alfvén wavepackets) propagating in opposite directions (see equ.(15.1)), the coherent interaction time is reduced to τ_A, which is smaller than the eddy-turnover-time by the factor $\mathrm{b}/\mathrm{V_a}$. The amplitude change during a typical "collision" is proportional to $\mathrm{b}/\mathrm{V_a}$. Assuming that successive collisions of wavepackets are independent, it is found that the time τ^* for a full interaction is such that $\tau_\mathrm{A} : \tau_{nl} : \tau^*$ or also: $\tau^* \approx (\tau_{nl}/\tau_\mathrm{A})\tau_{nl} \approx (\mathrm{V_a}/\mathrm{b})\tau_{nl}$, which may be much longer than τ_{nl}. Now, replacing the turnover-time by the effective transfer time τ^* in the expression of the energy transfer rate Π of equ.(2) , we obtain

$$\Pi_\mathrm{k} \approx \mathrm{k}\mathrm{E}^2/\mathrm{V_a} = \mathrm{k}^3\mathrm{E}_\mathrm{k}^2/\mathrm{V_a} \qquad (15.4)$$

and again imposing a constant dissipation rate ϵ, this leads to $\mathrm{E}(l) \approx (\epsilon \mathrm{V_a} l)^{1/2}$, or $\mathrm{E}_\mathrm{k} \approx (\epsilon \mathrm{V_a})^{1/2}\mathrm{k}^{-3/2}$, which is the IK spectrum.

There are however several factors which can prevent the IK decorrelation effect from taking place. First, the cascade process is not isotropic in wavenumber space, in presence of a large-scale magnetic field B°: for modes with wave-vectors nearly perpendicular to B°, the Alfvén decorrelation time becomes much larger than the nonlinear time (Shebalin et al, 1983; Grappin, 1986; Carbone and Veltri, 1990). Second, fluctuations are not purely incompressible, and this should lead to direct interactions between compressive waves propagating in the same directions (see e.g. Marsch and Mangeney, 1987).

Although the solar wind is neither incompressible, nor isotropic or homogeneous, the observed slopes are close to those inferred by the above arguments. While Coleman (1968) argued in favor of an IK spectrum (-3/2 slope), most of the observational evidence seems to be for spectral slopes in the solar wind turbulence very near the Kolmogorov slope, as soon as the heliocentric distance is larger than about 1 AU (see for example Bavassano and Smith, 1986). The present state of numerical simulations does not allow one to discriminate between the two values: even when the conditions are favor-

able (an incompressible, homogenous fluid with large scale magnetic fields) it is difficult, because of the limited resolution available, to measure spectral slopes accurately enough, and thus to determine which of the Kolmogorov or IK phenomenology is valid (see, e.g., the two-dimensional simulations by Biskamp and Welter, 1989). The preceding arguments assume implicitly that both field amplitudes z^+ and z^- are comparable. As remarked before, in "Alfvénic " situations this is not the case. From the MHD equations (1a), one sees that the turnover times for z^+ and z^- eddies in reality depends on the amplitude of the other field:

$$\tau_{nl}^{\pm} = l/z^{\mp} \tag{15.5}$$

Nonlinear interactions conserve separately the energies $E^{\pm}$ in both fields in the incompressible limit (eqs.(1)). It is thus legitimate to consider the possibility of separate energy cascades, via distinct fluxes Π^+ and Π^-. If one assumes that the IK decorrelation effect holds, then equal Π^+ and Π^- fluxes are obtained:

$$\Pi_k^+ = \Pi_k^- = k^3 E_k^+ E_k^- / V_a. \tag{15.6}$$

This formula again leads to the IK spectrum (proportional to $k^{-3/2}$) when both amplitudes are comparable (Dobrowolny et al, 1980b). When the $z^{\pm}$ asymmetry becomes large, eq.(15.5) does not provide a unique determination of the spectral slopes m^+, m^-; it only predicts that their sum should be equal to 3 (Grappin et al, 1983):

$$m^+ + m^- = 3. \tag{15.7}$$

In order to obtain a relation between the slopes and the fluxes a more elaborate theory is needed: we shall come back to this point in the next section. On the other hand if one assumes the interactions to be coherent as in Kolmogorov's cascade (see Matthaeus et al. 1983), i.e., the transfer times $\tau^{*\pm}$ to be equal to the turnover times, then the transfer rates are distinct (Tu, 1988):

$$\Pi_k^{\pm} = k^{5/2} E_k^+ E_k^- / (E_k^{\mp})^{1/2} = \epsilon^{\pm}. \tag{15.8}$$

In this case, the constraint of constant fluxes leads to the Kolmogorov spectra for both fields, whatever the asymmetry in the fluxes $\epsilon^{\pm}$.

15.3.3 Nonlinear evolution in the spherically expanding solar wind.

The equations (15.1) are valid for a homogeneous incompressible and stationary medium (constant U and V_a). In the solar wind, the

expansion effects appear as supplementary terms in the equation, which involve the "average" Alfvén and advection velocities V_a and U and their radial derivatives. The definition of the timescale T over which the full equations are averaged in order to separate the large-scale wind expansion from the small-scale fluctuations is somewhat arbitrary. The basic requirement is that T^{-1} should be less than the lowest frequency considered. The amplitude can then be split in two parts, an average $< \mathbf{z} >$ and a fluctuating $\mathbf{z}$. Upon subtraction of the time average from the original MHD equations and assuming incompressible fluctuations, one obtains (Whang, 1980):

$$\frac{\partial \mathbf{z}^{\pm}}{\partial t} + (\mathbf{V}_g{}^{\pm}) \cdot \nabla \mathbf{z}^{\pm} + \mathbf{z}^{\mp} \cdot \nabla (\mathbf{V}_g{}^{\mp}) + \frac{1}{2}(\mathbf{z}^{\pm} - \mathbf{z}^{\mp})(\frac{1}{2}\nabla \cdot \mathbf{U} \pm \nabla \cdot \mathbf{V}_a)$$
$$= -\frac{1}{\rho}\nabla p^T - (\mathbf{z}^{\mp} \cdot \nabla \mathbf{z}^{\pm} - \langle \mathbf{z}^{\mp} \cdot \nabla \mathbf{z}^{\pm} \rangle), \tag{15.9}$$

where $\mathbf{V}_g{}^{\pm} = \mathbf{U} \mp \mathbf{V}_a$ is the linear group velocity . An equation for the evolution of the quadratic quantities, such as the energy spectrum, may be obtained by multiplying eqns. (15.2) by $\mathbf{z}^{\pm}(\mathbf{x}', t')$, averaging and then Fourier transforming with respect to the "fast" variables $\mathbf{x} - \mathbf{x}', t - t'$. Assuming isotropy and time stationarity of the spectral quantities, one obtains equations for the spectra which depend both on wavenumber k (fast variable) and distance $\mathbf{r}$ (slow variable) (Marsch and Tu 1989):

$$\nabla \cdot (\mathbf{V}_g^{\pm}\ E_{\mathbf{k}}^{\pm}(\mathbf{r})) + E_{\mathbf{k}}^{\pm}(\mathbf{r})\frac{1}{2}\nabla \cdot \mathbf{U} + M_{\mathbf{k}}^{\pm}(\mathbf{r}) = \frac{\partial \Pi_k^{\pm}}{\partial k}. \tag{15.10a}$$

where the derivatives in the lhs denote derivatives with respect to the slow variable r. In eq.(15.10a) the first term is the wave energy flux, the second term is the work done by the waves in accelerating the bulk flow. The third term $M^{\pm}$ represents the linear coupling between the two wave species, due to the large scale inhomogeneities:

$$M^{\pm}(\mathbf{k}, \mathbf{r}) = -\mathbf{z}_{\mathbf{k}}^{+}\mathbf{z}_{\mathbf{k}}^{-} :: \left(\nabla \mathbf{V}_g^{\pm} - \frac{1}{2}\delta_{ij}(\nabla \cdot \mathbf{V}_a \pm \frac{1}{2}\nabla \cdot \mathbf{U})\right) \tag{15.10b}$$

and effectively couples the $\pm$ equations with that for the residual energy $E^r = \mathbf{z}^+ \cdot \mathbf{z}^- = (u^2 - b^2)$. Its order of magnitude may be estimated as $\approx E^r/\tau_{ad}$. In the linear case of transverse Alfvén waves, $M^{\pm}$ and the nonlinear terms may be neglected,(the distinction between energy spectral densities and energies at a given scale may in this case be dropped without altering the equations), and eq.(15.10) reduces to the equation of conservation of the adiabatic invariant

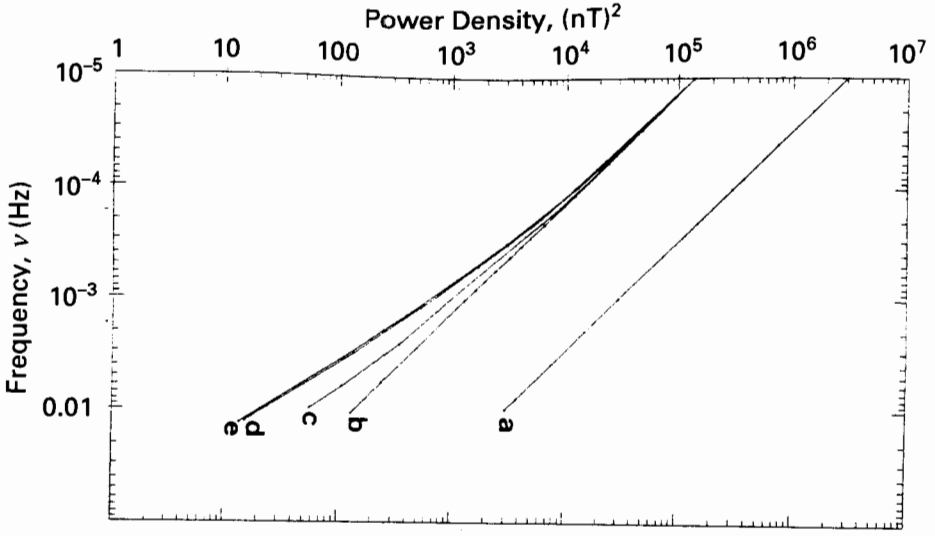


Figure 15.4. A schematic view of the evolution of energy spectra under the combined effects of wind expansion and nonlinear dissipation of turbulent energy in the model of Tu et al. 1984. (From Velli et al, 1990). a) initial spectrum at 0.3 AU b) WKB evolution at 1 AU c),d),e) evolution at 1 AU for increasing values of the constant A

$S^{\pm} = E^{\pm}/(k.V_a)$:

$$\nabla\cdot(\mathbf{V}_g^{\pm}S^{\pm}) = 0.$$

In the solar wind the Alfvén speed may be neglected compared to the wind speed, so that we obtain from equation (10) (Jacques, 1977):

$$\nabla\cdot(\mathbf{U}\rho E^{\pm}) + \rho E^{\pm}\nabla\cdot\mathbf{U}/2 = 0$$

which in a spherical expansion at constant speed gives $E^{\pm} \propto r^{-1}$ (recall that $E^{\pm}$ is the specific energy; the result usually quoted is $\rho E^{\pm} \propto r^{-3}$). Both spectra E^+ and E^- thus decrease with distance in a self-similar way, i.e., without changing their shape. The observations confirm that this simple linear description is valid at low frequencies, between $(2\cdot10^{-4}, 2\cdot10^{-3})$ Hz, while at larger frequencies they show that total turbulent energy decays more rapidly than simply predicted by the adiabatic change, suggesting again that turbulent dissipation (and thus nonlinear interactions) are at work (Bavassano et al, 1982; Schwenn 1983).

These considerations apply only in the inner heliosphere. Indeed, at larger distances from the Sun, the energy decay rate appears to be roughly frequency independent (Bavassano and Smith, 1986). The radial dependence, however, is $\propto r^{-}3.5$, which is slightly steeper than the WKB dependence.

In order to explain the changes in spectral shapes in the inner heliosphere, Tu et al (1984) and Tu (1988) developed a model in which the nonlinear flux on the rhs of eq.(15.10) is calculated following dimensional analysis as described in the previous section (equ.(15.6) or equ.(15.8): isotropy in $\mathbf{k}$ is necessarily assumed, so that the flux depends on the wavevector amplitude only). To close their equations, they assume the residual energy to be negligible, ($M^{\pm} = 0$). Moreover, the ratio $\alpha = E^+/E^-$ is chosen to be a constant, fitting an average observed value, and a spherically symmetric background wind is assumed, so that they finally obtain from eq.(15.10) a single equation for the spectral evolution of E^+, with the nonlinear flux term given by one of the two expressions:

$$\Pi_k^+ = A\frac{\alpha k^3 E_k^+(r)^2}{V_a} \quad (a), \qquad \Pi_k^+ = B\alpha k^{5/2} E_k^+(r)^{3/2} \quad (b), \qquad (15.11)$$

for the cases of strong magnetic fields (IK) and Kolmogorov's theory respectively (A and B are constants of order unity). The model describes the two regimes on either side of the critical scale $L(r)$ at which $\tau_{nl} \approx \tau_{ad}$. At small wavevector the nonlinear effects are negligibly small, the only modifications are limited to the WKB decay $u^2 \propto r^{-1}$; at larger wavevector they dominate so that locally an equilibrium spectrum in $k^{-5/3}$ is maintained (or $k^{-3/2}$, depending on the form (a) or (b) of the nonlinear flux (11)) the WKB effects acting to fix the energy level at a given frequency (see Figure 15.4). In this way, when starting with a flat (k^{-1}) spectrum as similar to what is observed at 0.3 AU, one obtains the steepening towards equilibrium spectrum, first at high frequencies, and later on at low frequencies.

15.4. Is "Alfvénic" turbulence different from "standard" MHD turbulence?

The model by Tu et al (1984) and Tu (1988) approximates the ratio E^+/E^- by an observational constant. However, both E^+ and E^- obey coupled evolution equations, the coupling occurring through expansion and nonlinear effects. Hence there is no reason to expect that the ratio remains constant. Then one may speculate on whether the situation found in the Alfvénic periods, where $z^- \ll z^+$, is the result of the evolution of the turbulence, or is a distinct property of high speed winds.

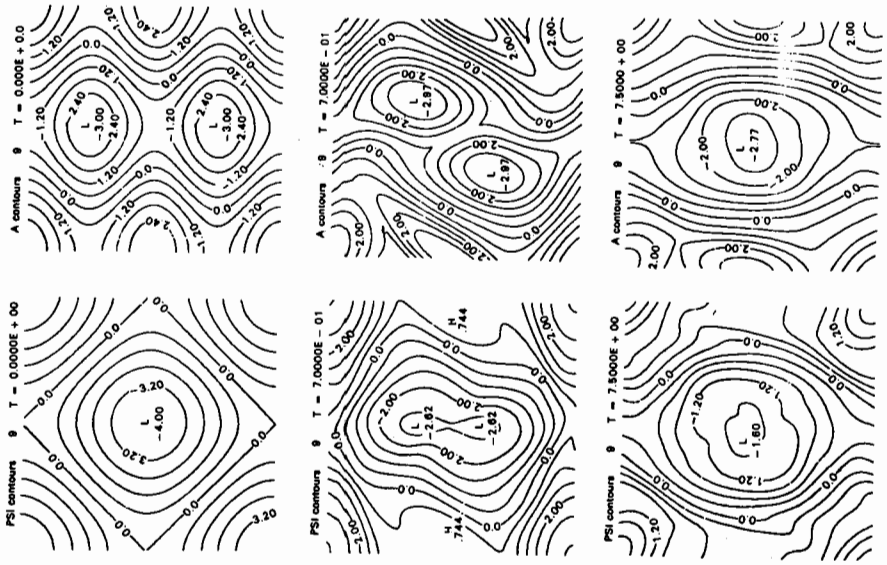


Figure 15.5. A two-dimensional MHD incompressible simulation showing the evolution of velocity (left) and magnetic field (right) towards a more and more aligned state. Time grows from top to bottom. (From Images de la Physique, CNRS, Paris, 1980 and Pouquet et al.1986).

15.4.1 Outward/inward evolution

A particular MHD mechanism, called "dynamic alignment" was proposed by Dobrowolny et al (1980b) to explain why the spectra were well-developed even when the inward mode energy E^- is very small. They suggested that in MHD turbulence, the velocity and magnetic field become more and more aligned due to nonlinear interactions, thus enhancing with time any initial imbalance between both inward and ouward component. The argument, which follows directly from the IK theory, is that the turbulent dissipation of both energies is the same (equ.(15.6)), so that the energy difference, i.e. the v.b correlation $C = (E^+ - E^-)/2$ stays constant while the total energy $E = (E^+ + E^-)/2$ decays, and the proportion $\gamma = C/E$ of the initially dominant (z^+) species indeed grows with time during the nonlinear interactions.

Numerical calculations have since fully confirmed the alignment process in homogeneous isotropic turbulence; in fact, the correlation C decreases, but less rapidly than energy, and this makes γ grow, i.e. the velocity and magnetic fields become more and more aligned (see Grappin et al., 1982; Matthaeus et al., 1983; Grappin, 1986; Pouquet et al, 1986) as shown in Figure 5 taken from Pouquet et al (1986).

The alignment process is obtained for a wide range of initial conditions, but it should be kept in mind that there are also other possible evolutionary tracks for the evolution of MHD turbulence, see Ting et al (1986). However, the evolution towards alignment is very slow, in fact too slow to take place in the solar wind between, say, 0.3 and 1 AU (Grappin et al., 1982). Second, and most importantly, the evolution actually seems to proceed in the reverse way in the solar wind: Roberts et al. (1987) found that in the average, the outward dominance tends to decrease rather than increase with heliocentric distance in the 0.3 AU-20 AU range. The explanation of the decrease of outward dominance at long distances may be that, in the solar wind, most of the energy is initially stored in the velocity shears due to the coexistence of high-speed and low-speed streams (Sturrock and Hartle, 1966): the nonlinear transfer of this pure kinetic energy towards small scales should sweep out any initial outward/inward imbalance of smaller scale eddies (see the closure calculations by Grappin et al. 1983, the simulations by Roberts and Goldstein (1988), and the analysis by Roberts et al. (1989)).

There is still another, purely linear way to couple z^+ and z^-. This coupling occurs because of the expansion, and is significant when the residual energy is not zero (term $M^\pm \approx E^r/\tau_{ad}$ in eq.(15.10)). Zhou and Matthaeus (1989) stressed that this term, being nonzero in the Alfvén range of the solar wind turbulent spectrum, could have important effects on the radial evolution of the energies $E^\pm$. Assuming isotropy, they obtain

$$M^\pm = D(u^2 - b^2)\nabla\cdot\mathbf{U}$$

where D is a constant, which we find equal to 1/6, instead of the value D=3/28 given in the original paper. Further, neglecting the Alfvén speed compared to the bulk speed, and the nonlinear terms of eq.(15.10), they obtain coupled linear equations for E^+, E^- (we don't write the third equation for E^r; see Marsch and Tu (1989) for a complete and rigorous derivation):

$$\mathbf{U}\cdot\nabla E^\pm + \nabla\cdot\mathbf{U}/2E^\pm = -D(u^2 - b^2)\nabla\cdot\mathbf{U} \qquad (15.12)$$

Now, instead of solving eq.(15.11) and the third equation for the residual energy $E^r = u^2 - b^2$, the authors use an empirical relation found in the "average" solar wind (see Figure 15.3):

$$\frac{(u^2 - b^2)}{(u^2 + b^2)} \approx -0.5, \qquad (15.13)$$

which allows them to eliminate the residual energy and thus to close the equations (15.12). The result is that the proportion of outward waves decreases with distance as observed. However, this agreement may be fortuitous, since the predicted energy decay with distance is much slower than observed. Zhou and Matthaeus (1989) emphasized that although nonlinear terms are neglected in this argument, the basic ingredient which modifies the WKB evolution is hidden in the observational constraint given in eq.(15.13) which can result only from nonlinear processes. The isotropy assumption is also crucial for their model: the opposite sign in front of the rhs of eq.(15.13), and thus the reverse result for the variation of outward/inward ratio would be obtained if one neglected the $\mathbf{z}^{\pm} \cdot \nabla \mathbf{U}$ terms in eq.(15.9), as would be natural for planar polarized waves for instance. Whatever the relevance of this argument for the observed decrease of the z^+/z^- ratio in the solar wind, the interesting point remains that, via the linear coupling between both fields due to the expansion, the respective evolutions of both $E^{\pm}$ energies must depend on their spectra in a complicated way.

Can we explain the observed level of residual energy? As we already said, the rhs of (15.12) is zero for pure Alfvén waves. In homogeneous MHD turbulence, Pouquet et al. (1976) found with closure methods that the residual spectrum was negative but decreasing faster than the energy spectrum (as k^{-2}), which has recently been found compatible with direct simulation results (Biskamp and Welter 1989). This result may be explained as follows (Grappin et al, 1983). As shown by Pouquet et al. (1976), the residual energy $E^r = u^2 - b^2$ is subject to a linear damping term, which arises from the phase mixing in the statistically isotropic magnetic field of the large-scale eddies. The phase mixing occurs within a time scale of the order of the Alfvén time τ_A, and acts as a sink for the residual energy which is balanced by a source provided by the nonlinear terms E_k/τ^*:

$$\frac{dE_k^r}{dt} \approx \frac{-E_k^r}{\tau_A} + \frac{E_k}{\tau^*}.$$

This equation has the interesting consequence that, in the almost symmetric case, $E^+ \approx E^- \approx E$, the residual energy spectrum is forced to follow the energy E, $E_k^r \approx \tau_A E_k/\tau^*$ in an Alfvén time, i.e. in a time much shorter than the energy transfer time τ^*. At equilibrium, the spectral transfer rate $\Pi_k \approx kE_k/\tau^*$ is scale-independent, and thus the residual spectrum scales as $\tau_A/k \propto k^{-2}$, which implies

that E_k^r/E_k is not constant, but decreases with increasing k. This argument provides some support to the idea underlying the use of a relation like (15.13), although the detailed form of the scaling may be different. One should also remember that in the solar wind these simple scalings could be modified by the other terms related to the expansion.

15.4.2 Turbulence is different when the outward mode strongly dominates: theoretical considerations

Let us now come to situations in which the outward component dominates strongly. The slope of the spectrum, in these situations, may become very different from the canonical slopes predicted by the standard phenomenologies. In numerical models of homogeneous and isotropic turbulence (derived from the primitive equations via the eddy-damped quasi-normal markovian (EDQNM) approximation), it is observed that the spectral slope of the dominant component (say, z^+) grows with the E^+/E^- ratio while that of the dominated one decreases, in agreement with the relation (15.7) (Grappin et al. 1983). This is somewhat supported by direct numerical simulations (Pouquet et al 1988; Politano et al 1989). It is necessary to invoke the contribution of nonlocal interactions in k-space to suppress the indeterminacy in eq.(15.6); this leads to the following relation between slopes and fluxes: $m^+/m^- \approx \Pi^+/\Pi^-$, which satisfactorily describes the EDQNM numerical results (Grappin et al, 1983).

In the solar wind, one observes (as we shall see in more detail in section 15.4.3) a strong dependance of the spectral shapes on the value of E^+/E^- but this relation is clearly much more complex than simulations of homogeneous turbulence would indicate. Again, the expansion effects must be taken into account when considering the solar wind. An important point is that in an inhomogeneous medium, both z^+ and z^- are a mixture of waves propagating in opposite directions. There is indeed in the WKB solution for each field $z^{\pm}$, as has been stressed by Velli et al.(1989), a component which propagates in the direction opposite to that found in the homogeneous case (see also Heinemann and Olbert, 1980). Some consequences of this are that:

1. The amplitude of the z^+ (say) field which is observed measures neither the amplitude of the outward component, nor that of the inward component, but a mixture of both. However, if $\eta = \tau_A \nabla \cdot \mathbf{U} \approx \tau_A/\tau_{ad}$ is smaller than unity (i.e. if the ratio l/R is smaller than

$V_a/U \approx 0.1$, which is true for periods T smaller than about 2 hours at $r \approx 0.3$AU), the contribution of the abnormal component in the z^+ species is of order η, so that the classical interpretation of z^+ and z^- fields remains meaningful, though approximate.

2. The z^- component can be entirely made of its abnormal component. This is precisely the case when there are only outward propagating waves, in which case the z^-/z^+ ratio (zero in the absence of expansion), is now about $\eta = \tau_A \nabla \cdot \mathbf{U} \approx \tau_A/\tau_{tr} \approx (l/R)(U/V_a)$. Thus, purely unidirectional waves can interact nonlinearly, since z^- is not zero. This nonlinear coupling will also create a standard component for the z^- fields, but one may show that there is a range of scales in which the non-standard couplings dominate (Velli et al, 1990). In these scales one therefore expects a cascade, and hence a spectral law, different from standard. To investigate the characteristics of this non-standard cascade, instead of fixing the z^-/z^+ ratio to a constant value, assume z^- to be equal to the abnormal component calculated from the dominant z^+ in the linear WKB approximation. One then obtains a stationary spectrum in k^{-1}. Velli et al. (1989) proposed that these arguments might explain the presence of well-developed k^{-1}spectra in regions with large z^+/z^- dominance at about 0.3 AU. However, the main problem with this theory is still that the nonlinear time necessary to build the spectrum becomes too large when z^- is too small (see the criticism of the preceding paper by Zhou et al. (1990) and subsequent reply, but also Velli et al. 1990).

15.4.3 Turbulence is different when the outward mode strongly dominates: observations

In their first observations of "Alfvénic periods" in the solar wind, Belcher and Davis (1971) remarked that the dominance of outward component was associated with high-speed streams.

More generally, the properties of turbulence vary with the global parameters of the plasma. For instance, 1-hour average data show a remarkable correlation between the magnetic rms fluctuation level (calculated on the same interval) and the average magnetic field (see Roberts et al., 1990). It has been conjectured that a plasma instability could lead to the saturation of the fluctuation level so as to maintain an approximately constant value of the ratio $\delta B/B$ (Hollweg, 1975).

However, this does not hold at higher frequencies, where a higher fluctuation level is found in high-speed than in low-speed streams;

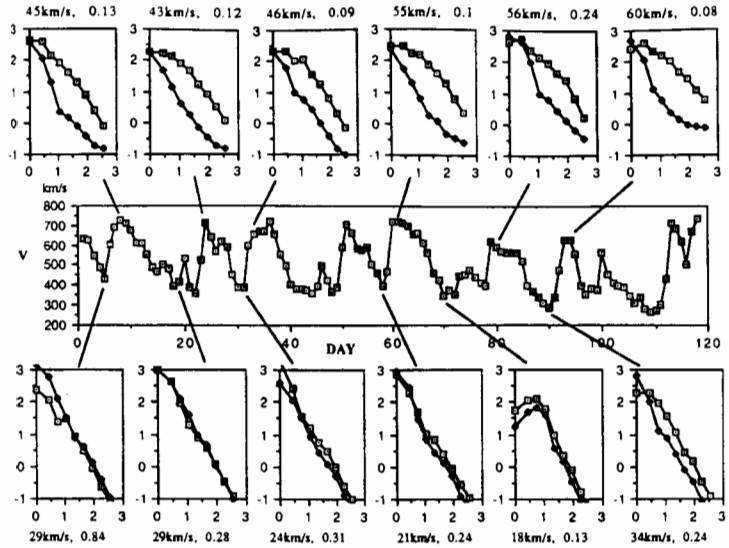


Figure 15.6. Some sample spectra in high-speed streams and interaction regions (First four months of Helios 1 mission, near solar minimum). Abscissa is time in days and ordinate is average bulk velocity. The small panels show energy spectra for the outgoing component (squares) and for the ingoing component (normalized to 1 $(km/s)^2$ day = $8.6 \cdot 10^4$ $(km/s)^2$ Hz^{-1}; frequency is normalized to $f_1 = 1$ day^{-1}). Numbers above top panels and under bottom panels indicate the proton thermal velocity c_s and the rms relative density fluctuation Δ n/n. (From Grappin et al. 1990a).

this suggests that the spectral slope depends on the large-scale properties of the wind (Denskat and Rosenbauer, 1982).

In order to investigate this point, Grappin et al (1989, 1990a) have reconsidered Helios data for a 4-month period during solar minimum activity and have made a detailed analysis of the variation of E^+ and E^- spectra as a function of the plasma parameters (see also Tu et al, 1989). Figure 15.6 taken from Grappin et al. (1990a) shows a series of sample z^+ and z^- spectra computed on a daily basis. The samples are taken respectively in peaks of velocity (top panel) and at speed minima, just in front of the peaks (bottom panel). The figure also shows the relationship with the bulk velocity pattern (middle panel). It shows a large variation of the spectra with the stream structure, in agreement with the preceding observations by Denskat et al (1981).

Note that a very similar behaviour has been found in a period near solar maximum (Grappin et al 1990b), which suggests that the results found in the solar minimum period are not peculiar, contrary to what has been argued by Roberts et al (1990).

In figure 15.6, one notes a kind of "breathing" which makes the

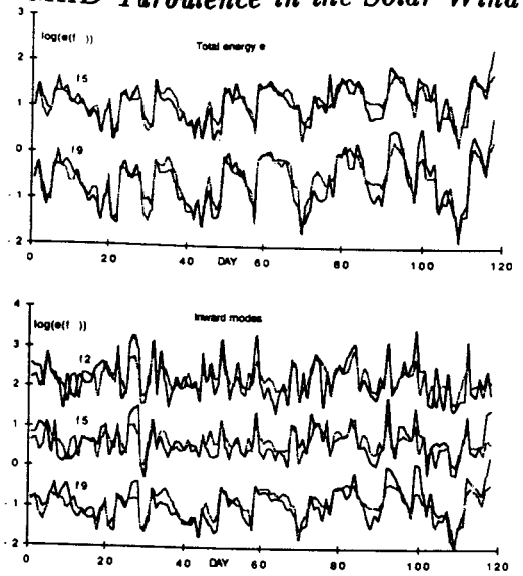


Figure 15.7. Helios 1 mission as in Figure 6. Comparison of the data with the prediction of a logarithmic fit with c_s and $\Delta n/n$. (a) total energy (kinetic plus magnetic) in bands 5 and 9 (fit with c_s only). (b) inward component's energy in bands 2, 5 and 9 (fit with both c_s and $\Delta n/n$). Solid line indicates original data, and dashed line is result of the fit. The frequency bands are octaves from $f_1 = 1$ day^{-1}: band n is $[2^{n-1}f_1, 2^n f_1]$. (From Grappin, Mangeney, Marsch 1990a).

ouward and inward components follow the velocity pattern, but in opposite ways. Grappin et al (1990a) remarked that the turbulent energy is in fact better correlated with proton thermal speed c_s and relative rms density fluctuation $\Delta n/n$ than with the bulk velocity. The relation is frequency dependent: for instance, high frequencies depend much more on thermal speed than low frequencies, and the reverse for $\Delta n/n$. The result of the fit by both parameters in several frequency bands is illustrated in Figure 7.

Note that the average proton thermal speed and Alfvén velocity are themselves rather well correlated, so that these results and those of Roberts et al (1990) mentioned above are compatible, taking into account the fact that the turbulent fields considered are different. However, the slope variation excludes that $\delta B/B$ is stream-independent at scales smaller than one hour.

A remarkable fact is that in the Alfvén range the spectra may well be approximated by power laws, with both the level and the slope of the spectra depending on the thermal speed and level of density fluctuation. The outward (resp. inward) component's energy level varies with temperature as c_s^2 (resp. c_s) at the lowest frequencies

and c_s^4 (resp. $c_s^{2.5}$) at the highest frequencies. The inward component varies as $\Delta n/n$ at low frequencies, and is insensitive to $\Delta n/n$ at high frequencies while the outward component is almost independent of $\Delta n/n$ in the whole frequency band. The spectral slopes for outward ($E^+(f) \propto f^{m+}$) and inward ($E^-(f) \propto f^{m-}$) components vary finally as:

$$m^+ \approx -1.67 + \log(c_s/c^*), \qquad (15.14a)$$

$$m^- \approx m^+ + 0.2\log(c_s/c^*) - 0.6\log(\frac{\Delta n/n}{\delta^*}), \qquad (15.14b)$$

where $c^* = 21.5$ km/s, $\delta^* \approx 0.28$ are reference values, which are those typically found in the slow wind, characterized by low temperature and large density fluctuations; (these density fluctuations may be interpreted as slow magnetosonic waves, or pressure-balanced convected structures (see Bavassano et al, 1989)). With the reference values $c_s = c^*$, and $\Delta n/n = \delta^*$ equ. (15.13) gives the Kolmogorov value for the spectra of both outward and inward components. In this case, the observations and the fit also show that both the outward and inward energies are about equal. In the high speed flows on the contrary, the proton thermal speed may reach values as high as 60km/s and the density fluctuations fall down to minimum values about $\Delta n/n \approx 0.06$. When we use these figures in eq.(15.14), we obtain spectral slopes much flatter than Kolmogorov's -5/3: $m^+ \approx -1.2$, $m^- \approx -0.8$, while the level of outward component is much higher, and that of the inward component much lower (due to the fall of $\Delta n/n$), compared to the cold wind conditions: these are the Alfvénic periods. Therefore it appears that:

A. There is no abrupt transition between the two kinds of plasmas ("Alfvénic" and "non-Alfvénic"), but a continous family of intermediate states, parametrized by both proton temperature and amplitude of relative density fluctuations $\Delta n/n$.

B. Once the large, systematic variation with distance of the average proton density is eliminated by considering energy per unit mass instead of per unit volume, the radial variation of turbulent energy and slopes is negligible (recall that in the Helios data the radial distance varies from 0.3 to 1 AU) compared to the slow time variation (slow meaning here variations of a one day average in the spacecraft frame) associated with temperature and density variations. For example, the energy decrease in r^{-1} due to pure WKB effects corresponds to a factor 3 variation between 0.3 and 1 AU, while the daily variation of the spectra covers several orders of magnitude. Accord-

ingly, the average radial temperature variation is much smaller than the daily temperature variation (Freeman, 1988). A possible interpretation is that the fluctuations of the turbulent state of the wind observed from day to day in the satellite's records reflect essentially the variations in the solar parameters as they may be found at the source of the wind. We would thus have at the source a whole range of turbulent spectra between the two extremes, namely the equilibrium Kolmogorov spectra, and the flat Alfvénic spectra. This view is not incompatible with a nonlinear relaxation phenomenon such as proposed by Tu et al. (1984) and Tu (1988) which would lead everywhere at long distances to a "Kolmogorov" turbulence. To ascertain this point, it would be necessary to investigate the variations of the outward and inward spectra at large distances from the Sun, for example in the Voyager data.

What can we say now, using the detailed observations of both $z^{\pm}$ components, on the problem of heating the high-speed streams? We noticed at the beginning of this review that, there was no major problem to explain the heat input from turbulent dissipation by the waves in slow flows, using simple dimensional estimations of nonlinear coupling terms for the waves (Hollweg 1986). On the other hand, as mentioned by Belcher and Davis (1971), the hot, high-speed streams contain precisely the "Alfvénic" turbulence which, as we have emphasized earlier, contain smaller nonlinear couplings. This apparent contradiction can actually be solved, if we take into account the respective levels of turbulence in fast and slow winds (see Grappin et al, 1990a), and distinguish between outward and inward energy fluxes and timescales. Consider indeed the outward energy flux Π^+ through the scale l:

$$\Pi^+ \approx z^- z^{+2}/l;$$

the ratio between the heating rates per unit distance in an Alfvénic (fast wind) period and that in a "slow wind" period is

$$(\Pi_f^+/U_f)/(\Pi_s^+/U_s) \sim 1.5$$

since

$$\Pi_f^+/\Pi_s^+ \approx (z^- z^{+2})_f/(z^- z^{+2})_s = (z_f^+/z_s)^3 (z^-/z^+)_f \approx 3,$$

if we assume, as the observations show (see for instance figure 15.6), that z_f^-, $z_s^+ \approx z_s^- \approx z_s$, and z_t^+ are ordered following a growing geometrical progression of a factor about 3. This larger flux is not contradictory with a slower evolution of the spectra:

$$\tau_s^+/\tau_f^+ \approx (z_f^-/z_s^-) \approx 0.3.$$

15.5. Conclusion

From the astrophysical point of view, one may say that the main thing that we have learnt is how MHD fluctuations can heat the wind even at some distance from the stellar surface. Accordingly, the observed temperature decrease with distance is slower than the adiabatic one. This is true whether the turbulence has or has not (as in the high-speed streams) relaxed to an equilibrium Kolmogorov-like state. The problem of the existence of the high-speed streams themselves remains however unsolved.

Let us now leave the domain of astrophysics for that of fluid mechanics. The solar wind has long been considered as a potential laboratory for MHD turbulence. A specific physical effect such as dynamic alignment of velocity and magnetic fields has been found while attempting to answer problems posed by the analysis of solar wind data.

One of the arguments which has been most widely used in the interpretation of the solar wind data, based on the standard dimensional analyses, is that the natural outcome of the non-linear cascade is a universal spectrum. However, numerical simulations have shown that these simple phenomenological models apply only when the turbulence is roughly symmetric. When this condition is violated the asymptotic spectra are not unique; this is also true in the solar wind, where the observed spectra are not universal, and also the degree of asymmetry may be quite high. In the numerical simulations of homogeneous MHD, the spectral shapes depend directly on the degree of asymmetry, and both change with time, more and more slowly, while the asymmetry increases. In the solar wind, due probably to the fact that large-scale fluctuations are essentially kinetic, the $z^{\pm}$ asymmetry actually decreases with time (i.e. with distance), which makes the large distance turbulence more akin to standard symmetric MHD. Second, the relation between asymmetry and spectral slope is not only completely different from the theoretical one, but moreover, surprisingly, the parameters which appear to best organize the data are the proton temperature and the rms density fluctuation.

In particular, it is difficult to understand why the relative density fluctuations are an order of magnitude lower in the Alfvénic periods

than in the non-Alfvénic periods, while the turbulent Mach numbers are comparable or higher. This contradicts the known fact that a compressible flow naturally develops density fluctuations whose amplitude depends essentially on the turbulent Mach number (see Lighthill, 1967, Passot and Pouquet 1987 for numerical experiments in the hydrodynamic case, and Montgomery et al., 1987, Shebalin and Montgomery, 1988 in the MHD case). One often invokes the larger damping of magnetosonic waves to explain the low level of density fluctuations, in the Alfvénic periods, but then why is it that the same damping does not work in the slow winds? Or is it because there is not enough time to couple compressible and solenoidal fluctuations in the fast wind?

All this suggests that an interesting numerical experiment would be to simulate the evolution of turbulence in an expanding box. It should be possible to retain within reasonable computer costs a number of interesting properties:

- a supersonic turbulence
- the essential effects of the expansion, i.e., a term like $M^{\pm}$ in equation (15.3).

On the other hand, to obtain well-developed turbulent spectra would require very large simulations, which presently seems to be out of reach. However, we should be able to answer in a quantitative way to questions like:

- what is the effect of the residual energy: is it or not enslaved by the other quantities?
- is there a specific turbulent regime in an expanding MHD fluid dominated by purely outward-propagating waves?
- in this case, are the density fluctuations reduced, relative to the standard Lighthill-like sound production?

Acknowledgments

We thank D.A. Roberts for careful reading the manuscript, and interesting criticisms and suggestions.

References

Alazraki, G., P. Couturier (1971), *Astron. Astrophys.* **13**, 380.

Barnes, A. (1979), Hydromagnetic waves and turbulence in the solar wind, in: Solar system plasma physics, vol. 1, (eds E. N. Parker, C. F. Kennel, and L. J. Lanzerotti), pp. 249-319, North-Holland, Amsterdam.

Barnes, A. (1983), Hydromagnetic waves, turbulence and collisionless process es in the interplanetary medium, in Solar-Terrestrial Physics, (ed. by R.L. Carovillano and J.M. Forbes), D. Reidel, Dordrecht, P.155,.

Bavassano,B. and R. Bruno (1989), Large-scale solar wind fluctuations in the inner heliosphere at low solar activity, *J. Geophys. Res.* **94**, 168.

Bavassano, B., M. Dobrowolny, F. Mariani, and N. F. Ness (1982), Radial evolution of power spectra of interplanetary Alfvénic turbulence, *J. Geophys. Res.* **87**, 361.

Bavassano B., and E.J., Smith (1986), Radial variation of interplanetary Alfvénic fluctuations: Pioneer 10 and 11 observations between 1 and 5 AU, *J. Geophys. Res.* **91**, 1706.

Belcher, J. W., and L. Davis (1971), Large-amplitude Alfvén waves in the interplanetary medium, 2, *J. Geophys. Res.* **76**, 3534.

Biskamp D., and H. Welter (1989), *Phys. Fluids* **B1**, 1964.

Burlaga, L. F., (1987), Interaction regions in the distant solar wind, Proc. Sixth Int. Solar Wind Conf., Solar Wind 6, P.547, (ed. V. J. Pizzo, T. Holzer, D.G. Sime).

Carbone V., P-L. Veltri (1990), A shell model for anisotropic magnetohydrodynamic turbulence, *Geophys. Astrophys. Fluid Dyn.* **52**, 153.

Coleman, P. J. (1968), Turbulence, viscosity, and dissipation in the solar w ind plasma, *Astrophys. J.* **153**, 371.

Denskat, K. U., F. M. Neubauer, (1982), Statistical properties of low-frequency magnetic field fluctuations in the solar wind from 0.29 to 1.0 AU during solar minimum conditions: Helios 1 and Helios 2, *J. Geophys. Res.* **87**, 2215.

Denskat, K. U., F. M. Neubauer, (1983), Observations of hydromagnetic turbulence in the solar wind, Solar Wind Five, NASA Conf. Publ. CP-2280, P.81.

Dobrowolny, M., A. Mangeney, and P.-L. Veltri (1980a), Properties of MHD turbulence in the solar wind, *Astron. Astrophys.* **83**, 26.

Dobrowolny, M., A. Mangeney, and P.-L. Veltri (1980b), Fully developed anisotropic turbulence in the solar wind, *Phys. Rev. Letts.* **45**, 144.

Freeman, J. W., Estimates of solar wind heating inside 0.3 AU (1988), *Geophys. Res. Letts.* **15**, 88.

Grappin, R. (1986), Onset and decay of two-dimensional magnetohydrodynamic turbulence with velocity-magnetic field correlation, *Phys. Fluids* **29**, 2433.

Grappin, R., U. Frisch, J. Léorat, and A. Pouquet (1982), Alfvénic fluctuations as asymptotic states of MHD turbulence, *Astron. Astrophys.* **105**, 6.

Grappin, R., A. Pouquet, and J. Léorat (1983), Dependence of MHD turbulence spectra on the velocity field-magnetic field correlation, *Astron. Astrophys.* **126**, 51.

Grappin, R., A. Mangeney, and E. Marsch (1989), On the origin of solar wind turbulence: Helios data revisited, in Proceedings of the workshop on Turbulence and Nonlinear Dynamics in MHD Flows, (edited by M. Meneguzzi, A. Pouquet, and P. L. Sulem), p. 81, North-Holland, Amsterdam.

Grappin, R., A. Mangeney, and E. Marsch (1990a), On the origin of solar wind turbulence: Helios data revisited, *J. Geophys. Res.* **95**, 8197.

Grappin, R., A. Mangeney, and E. Marsch (1990b), MHD turbulence in the solar wind in Helios data: solar maximum versus solar minimum, in preparation.

Grappin, R., M. Velli and A. Mangeney, (1990c), in preparation.

Heinemann, M., and S. Olbert (1980), Non WKB waves in the solar wind, *J. Geophys. Res.* **85**, 1311.

Hollweg, J. V. (1975), Waves and instabilities in the solar wind, *Rev. of Geophysics* **13**, 263.

Hollweg, J. V. (1978), Alfvén waves in the solar atmosphere, *Solar Phys.* **56**, 305.

Hollweg, J. V. (1986), Transition region, corona, and solar wind in coronal holes, *J. Geophys. Res.* **91**, 4111.

Isenberg, P. A. (1989),The solar wind, Geomagnetism, vol.4, (ed. by J.A. Jacobs), Academic Press.

Iroshnikov, P. S. (1963), *Soviet Astron.***7**, 566.

Jacques, S.A. (1977), Momentum and energy transport by waves in the solar atmosphere and solar wind, *Astrophys. J.* **215**, 942.

Kolmogorov, A., N. (1941), *C. R. Acad. Sc. URSS***30**, 301.

Kraichnan, R. H. (1965), Inertial range spectrum of hydromagnetic turbulence , *Phys. Fluids* **8**, 1385.

Leer E., T.E., Holzer, and T. Fla (1982), Acceleration of the solar wind, *Space Sci. Rev.***33**, 161.

Leroy, B. (1981), Propagation of waves in an atmosphere in the presence of a magnetic field, *Astron. Astrophys.* **91**, 136.

Lighthill, M.J. (1967), Predictions on the velocity field coming from acoust ic noise and a gneralized turbulence in a layer overlying a convectively unstable atmospheric region, IAU Symposium 28, p.429.

Marsch, E. (1990a), Kinetic physics of the solar wind plasma, in Physics of the inner heliosphere, (ed. R. Schwenn and E. Marsch), Springer-Verlag, Heidelberg, in press.

Marsch, E. (1990b), MHD turbulence in the solar wind, in Physics of the inner heliosphere, (ed. R. Schwenn and E. Marsch), Springer-Verlag, Heidelberg, in press.

Marsch E., A. Mangeney (1987), Ideal MHD equations in terms of compressive Elsasser variables, *J. Geophys. Res.* **92**, 7363.

Marsch, E., C.Y. Tu (1989), Dynamics of correlation functions with Elsasser variables for inhomogeneous MHD turbulence, *J. Plasma Phys.* **41**, 479.

Matthaeus, W. H., M. L. Goldstein (1982), Measurements of the rugged invariants of magnetohydrodynamic turbulence in the solar wind, *J. Geophys. Res.* **87**, 6011.

Matthaeus, W. H., M. L. Goldstein, D. Montgomery (1983), Turbulent generation of outward traveling interplanetary Alfvénic fluctuations, *Phys. Rev. Letts.* **51**, 1484.

Matthaeus, W. H., M. L. Goldstein (1986), Low-frequency 1/f noise in the interplanetary magnetic field, *Phys. Rev. Letts.* **57**, 495.

Matthaeus, W. H., M. L. Goldstein and J.H. King (1986), An interplanetary magnetic field ensemble at one 1AU, *J. Geophys. Res.* **91**, 59.

Montgomery, D. (1983), Theory of hydromagnetic turbulence, Solar Wind Five, Nasa Conf Publ., CP 2280, 107.

Montgomery, D., M. R. Brown, and W. H. Matthaeus (1987), Density fluctuation spectra in magnetohydrodynamic turbulence, *J. Geophys. Res.* **92**, 282.

Parker E., N. (1963), "Interplanetary dynamical processes," Interscience, New York.

Passot T., and A. Pouquet (1987), Numerical simulation of compressible homogeneous flows in the turbulent regime, *J. Fluid Mech.* **181**, 441.

Politano, H., A. Pouquet, and P.L. Sulem (1989), Inertial ranges and resistive instabilities in two-dimensional magnetohydrodynamic turbulence, *Phys. Fluids* **B1**, 2330.

Pouquet, A., U. Frisch, and J. Léorat (1976), Strong MHD helical turbulence and the nonlinear dynamo effect, *J. Fluid Mech.* **77**, 321-354.

Pouquet, A., M. Meneguzzi, and U. Frisch (1986), Growth of correlations in MHD turbulence, *Phys. Rev.* **33A**, 4266.

Pouquet, A., P. L. Sulem and M. Meneguzzi (1988), Influence of velocity- magnetic field correlations on decaying magnetohydrodynamic turbulence with neutral X points, *Phys. Fluids* **31**, 2635.

Roberts, D. A., M. L. Goldstein, L. W. Klein, and W. H. Matthaeus (1987), Origin and evolution of fluctuations in the solar wind: Helios observations and Helios-Voyager comparisons, *J. Geophys. Res.* **92**, 12023.

Roberts, D. A., and M. L. Goldstein (1988), Simulation of interplanetary dynamical processes, in Proceedings of the Third International Conference on Supercomputing, (edited by L. P. Kartashev and S. I. Kartashev), p. 370, International Supercomputing Institute, Boston.

Roberts, D. A., M. L. Goldstein, W. H. Matthaeus, and L. W. Klein (1989), Observation and simulation of MHD turbulence in the solar wind, in Turbulence and nonlinear dynamics in MHD flows, (ed. by M. Meneguzzi, A. Pouquet and P.L. Sulem), North-Holland, Amsterdam.

Roberts, D. A., M. L. Goldstein, L. W. Klein (1990), The amplitudes of interplanetary fluctuations: stream structure, heliocentric distance and frequency dependence, *J. Geophys. Res.* **95**, 4203.

Schwenn, R. (1983), The "average" solar wind in the inner heliosphere: Structures and slow variations, Solar Wind 5, Nasa Conf. Publ., CP2280, 489.

Scott, S. L., W. A. Coles, and G. Bourgois (1983), Solar wind observations near the Sun using interplanetary scintillation, *Astron. Astrophys.* **123**, 207- 215.

Shebalin J.V., W.H. Matthaeus, D. Montgomery (1983), Anisotropy in MHD turbulence due to a mean magnetic field, *J. Plasma Phys.* **29**, 525.

Shebalin J.V. and D. Montgomery (1988), Turbulent magnetohydrodynamic density fluctuations, *J. Plasma Phys.* **39**, 339.

Sturrock P.A., Hartle, R.E. (1966), *Phys. Rev. Letts.* **16**, 628.

Ting A., W.H. Matthaeus, D. Montgomery (1986), Turbulent relaxation in magnetohydrodynamics, *Phys. Fluids* **29**, 3261.

Tu, C.(1988), The damping of interplanetary Alfvénic fluctuations and the heating of the solar wind, *J. Geophys. Res.* **93**, 7.

Tu, C.-Y., Z. Y. Pu, F. S. Wei (1984), The power spectrum of interplanetary Alfvénic fluctuations: Derivation of the governing equation and its solution, *J. Geophys. Res.* **89**, 9695.

Tu, C.-Y., E. Marsch, and K.M. Thieme (1989), Basic properties of solar wind MHD turbulence near 0.3 AU analyzed by means of Elsasser variables, *J. Geophys. Res.* **94**, 11739.

Velli, M., R., Grappin, A., Mangeney (1989), Turbulent cascade of unidirectional Alfvén waves in the interplanetary medium, *Phys. Rev. Letts.* **63**, 1807.

Velli, M., R. Grappin, A., Mangeney (1990), Solar Wind expansion effects on the evolution of hydromagnetic turbulence in the interplanetary medium, *Computer Phys. Comm.***59**, 153.

Whang Y., C. (1980), *J. Geophys. Res.* **85**, 2285.

Zhou Y. and W. H. Matthaeus (1989), Non WKB evolution of solar wind fluctuations: a turbulence modeling approach, *Geophys. Res. Letts.* **16**, 755.

Zhou Y. and W. H. Matthaeus, D.A. Roberts, M. Goldstein (1990), Physical consistency in modeling interplanetary MHD fluctuations *Phys. Rev. Letts.* **64**, 2592.

16

The Earth's Magnetosphere

MARK SAUNDERS

Imperial College, London SW7 2BZ, U. K

16.1 Introduction

Our understanding of the Earth's magnetosphere has developed steadily since the start of the space age and is now at a mature and sophisticated level. With this maturity we can appreciate just how fundamental magnetohydrodynamics or MHD is in ordering most of the key magnetospheric features. This knowledge has wider significance because similar MHD processes probably also occur on the Sun, in the Universe and in the plasma state; regions which, unlike the magnetosphere, have yet to be probed properly either in situ or remotely. The magnetosphere's importance is underlined further by the selection of SOHO (Solar Heliospheric Observatory) and CLUSTER (a multipoint magnetosphere mission) as the first objective in the E.S.A. Horizon 2000 programme of long-term space exploration.

I begin the chapter by placing the magnetosphere together with the reasons and techniques for studying it in context. Section two reviews the overall structure and characteristics of the magnetosphere, emphasising those regions and processes where MHD is fundamental. The third section focusses on the key advances (experimental and theoretical) in our exploration of magnetospheric MHD during the past few years, and the chapter concludes with a "look into the future".

16.1.1 Why study the magnetosphere?

The Earth's magnetosphere is the volume of space surrounding the Earth where all physical processes are controlled by the Earth's magnetic field (Gold, 1959). These processes involve the interplay of plasmas (having number densities varying between 10 cm^{-3} and 10^4 cm^{-3}) with magnetic fields (having strengths ranging from 10 nT to 10^4 nT); interactions which continuously redistribute energy and

momentum between flowing plasma populations and magnetic fields. Such interactions lead to plasma acceleration or deceleration and to plasma heating or cooling. This energy transfer can prove spectacular as demonstrated by the magnetic field energy release coming from magnetic reconnection at the dayside magnetopause where plasma flow speed increases of up to an order of magnitude may result. Most of the large-scale physics of the magnetosphere is described successfully by MHD and I shall emphasise these aspects of our knowledge in this chapter.

At the start of the space age in the late 1950s, the Earth's magnetosphere was studied simply for reasons of exploration: because "it was there". Now many countries are recognising that their national interests will benefit by increasing efforts to understand our space environment. The reasons for this are threefold. Firstly, magnetospheric physics is an essential element in the development of science; its impact extends into areas of solar physics, astrophysics, plasma physics, atomic physics and geophysics. Secondly, it has important practical applications, many of which will grow in relevance as humans develop advanced societies beyond the surface of planet Earth. Last but certainly not least, it offers a fine opportunity for forwarding international collaboration and detente.

The scientific areas for which magnetospheric physics has greatest significance are solar physics and astrophysics. These links are many and varied, for magnetised plasma systems are ubiquitous in the Universe and the processes occurring therein are believed to mimic those present in the Earth's magnetosphere. As "in situ" measurements are possible only within the magnetosphere, there is considerable scope for transferring knowledge and understanding between the "Earth" and, say, "Solar" plasma physics communities.

Understanding the behaviour of magnetospheric plasmas also has some important practical applications which affect operations both in space and here on Earth. The Earth's magnetosphere is a hostile environment to explore. It contains plasmas with temperatures as high as 10^8K and energetic particles with energies approaching 10^8eV. Not surprisingly, this environment proves dangerous to satellites and to astronauts alike. For example, a sudden increase in solar activity can seriously affect satellite orbit stability (witness the early loss of the Skylab mission), the performance of electronic components on satellites, Space Shuttle reentry times, and the survival of astronauts. With space now being used increasingly for various scientific, communication and defence studies, and with the future

prospect of the Space Station, transatmospheric air travel and human colonies in space, it is essential that space vehicles are able to function continuously in the near-Earth environment subject to the dynamic variations of the Sun and magnetosphere. Thus the predictions of 'space weather' may become as essential to future activities in space as the prediction of Earth weather is to our present daily lives.

16.1.2 Origins and Development of the Subject

Magnetospheric physics began as an identifiable discipline with the launch of the first Earth satellites in the late 1950's and the discovery in 1958 of the Van Allen radiation belts. However, its scientific heritage extends back much further and originates in the study of geomagnetism, notably the discovery in 1600 by Gilbert, physician to Queen Elizabeth 1 and a contemporary of Shakespeare, that "magnus magnes ipse est globus terrestris" (the Earth globe itself is a great magnet). A significant event in geomagnetism's history came in 1722 when Graham, an eminent London clock-maker, noticed that the magnetic field measured on the Earth's surface undergoes transient changes on time-scales of hours to days. The quiet daily magnetic changes eventually led to the recognition of the ionosphere by Balfour Stewart in 1882.

In addition to the regular daily variations, periods of magnetic disturbance or activity also occur. When a magnetic disturbance measured at the ground is considerable it is called a magnetic storm - a term introduced by von Humboldt in 1808. During the nineteenth century many theories were proposed to explain magnetic storms. However, the breakthrough came from the pioneering work of Chapman and Ferraro (1931, 1932, 1933) who proposed that magnetic storms result when clouds of solar plasma collide with the Earth's magnetic field (see Figure 16.1). A concept central to this theory was that the terrestrial magnetic field creates a vast cavity in the advancing solar plasma. Following the suggestion of Gold (1959), this 'geomagnetic cavity' became known as the magnetosphere. With the start of the space age the magnetosphere's complex and fascinating characteristics soon began to emerge. Many of these important characteristics will be described in Section 16.2 but first let us examine the research tools necessary for exploring the Earth's magnetosphere.

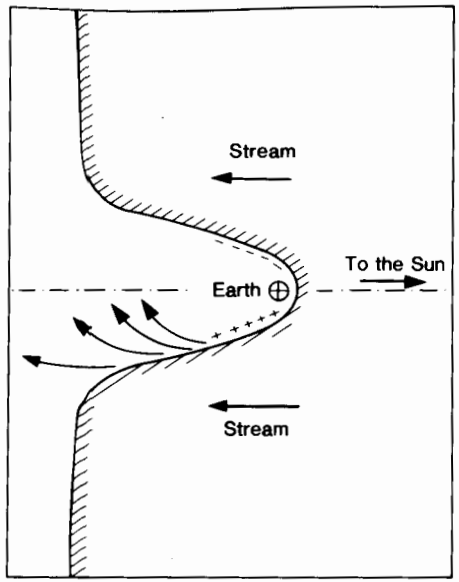


Figure 16.1. The Chapman and Ferraro (1931, 1932, 1933) model for a magnetic storm - the first model to hint at the existence of the Earth's magnetosphere. "Streams" or " clouds" of solar plasma compress the terrestrial magnetic field into a cavity. Some of this plasma enters the cavity and produces a ring of electric current surrounding the Earth. This "ring current" reduces the magnetic field recorded on the Earth's surface and is a primary indicator of a magnetic storm.

16.1.3 Research Techniques

The techniques used for studying MHD properties of the magnetosphere may be divided into four categories with the first two often simply being termed "experimentation": (a) "In situ" and remote sensing measurements by space probes and rockets; (b) Ground-based observations; (c) Data analysis; (d) Theory and simulation.

Current research in the field often has the following sequence: detailed experimental observations (especially "in situ" studies), data analysis, and theoretical interpretation.

Space probes provide the most powerful experimental tool for detailed studies of the magnetosphere. They carry an array of sophisticated plasma and electromagnetic field instruments capable of recording at various energies and over different frequency ranges. Ground-based observations come from a network of several hundred instruments distributed worldwide and from recordings made by sounding rocket flights. They are operated most commonly at

high magnetic latitudes, in the Arctic and Antarctic, to study questions of auroral physics and to further our knowledge of just how the magnetosphere and ionosphere are electrically coupled. Ground-based measurements are made by: (a) Magnetometers which record fluctuations in the Earth's magnetic field; (b) Backscatter radars which probe the ionosphere's response to magnetosphere forcing; (c) All-sky cameras and photometers which observe the light emissions emanating from the aurora airglow; (d) Radio receivers which monitor energetic particle precipitation.

There are few fields in physics where data analysis assumes the special status which it now holds in magnetospheric physics. With the advent of complex multi-instrument and multi-satellite data sets, essential for a complete understanding of the Earth's magnetosphere, data analysis has, perforce, established itself as an important technique in its own right. Indeed there are many scientists in the field who now regard themselves primarily as "data analysts". The work involved in data analysis is wide ranging. To analyse data and interpret it in terms of relevant theories requires familiarity with the operation and limitation of experiments, a sound theoretical knowledge and an ability to use computers. The inception of electronic communication networks, such as the SPAN (Space Physics Analysis Network) system in the US and Europe has greatly aided data analysis work.

Excluding data interpretation, theoretical methods may be divided into the two related categories of (a) analytical methods and (b) computer simulation methods. The analytical approach is the theorist's traditional tool and involves the use of a set of basic physical laws and equations for making precise theoretical predictions which may be tested. However, this procedure is unsuitable for studying large-scale global interactions whose complexity necessitates the use of numerical simulations on a computer. Such simulations provide the possibility for modelling magnetised plasma processes in three dimensions and at various time resolutions. There is little doubt that simulations will play an increasingly important role in future magnetospheric research.

16.2 An MHD Description of the Magnetosphere

The Earth's magnetosphere is a good example of an MHD system. In this section I review exactly what MHD can and cannot explain, describe the major electromagnetic and plasma regions of the mag-

netosphere and examine briefly the major MHD processes and phenomena.

16.2.1 When is MHD valid and what can it explain?

The MHD approximation treats plasma particles as behaving like a fluid with their properties determined by Maxwell's equations and by the equations of hydrodynamics. The generalised Ohm's Law

$$\mathbf{E} = -\mathbf{v} \times \mathbf{B} + \sigma^{-1}\mathbf{J} + \frac{1}{ne}\Big(\mathbf{J} \times \mathbf{B} - \nabla p_e\Big), \qquad (16.1)$$

illustrates when the MHD limit is valid and when it breaks down. This law states that the electric field E in a stationary (satellite) rest frame is the sum of four principal terms: (a) a motional ($\mathbf{v} \times \mathbf{B}$) electric field due to the plasma flow $\mathbf{v}$ carrying a magnetic field $\mathbf{B}$, (b) an ohmic resistance term $\sigma^{-1}\mathbf{J}$ due to electric current flow $\mathbf{J}$ through a plasma of resistivity σ^{-1}, (c) a Hall current term ($\mathbf{J} \times \mathbf{B}/ne$) due to the finite gyration time of the ions about $\mathbf{B}$, n being the plasma number density and e the electronic charge, and (d) an ambipolar electric field term which is required to retard the electrons in the presence of an electron pressure gradient p_e (thereby keeping the plasma neutral). The latter three terms correspond to static electric fields in the plasma itself and the MHD limit is where these terms are ignorable.

Dimensional analyses (see, for example, Siscoe (1983)) show that for most length l and time t scales within the magnetosphere, the motional electric field term easily dominates so that equation (16.1) reduces to

$$\mathbf{E} = -\mathbf{v} \times \mathbf{B} \qquad (16.2)$$

This is the MHD limit of the generalised Ohm's Law. The most important consequence of this limit is that the magnetic field is frozen into the plasma flow so that any two plasma particles on a magnetic field line will always remain on that same magnetic field line no matter how the flow develops. This "frozen-in-field" approximation provides great simplicity when interpreting magnetised plasma behaviour.

There is also a less rigorous but more workable definition for when MHD is valid. MHD is a good approximation provided the local magnetic field changes on scales large compared to the typical particle gyroradius and over intervals long compared to the typical particle gyroperiod, that is, when

$$\frac{r_L}{B}\frac{\partial B}{\partial l} \ll 1$$
$$\frac{\tau_c}{B}\frac{\partial B}{\partial t} \ll 1 \qquad (16.3)$$

where r_L and τ_c are the particle gyroradius and gyroperiod, B is the magnetic field strength and l and t are distance and time coordinates. Representative values for r_L and τ_c in the magnetosphere vary from 100 km and 1 sec in the outer magnetosphere to 10 m and 10^{-2} sec in the inner magnetosphere.

From the above discussion MHD is clearly a powerful tool for describing the large-scale structures and processes within the magnetosphere, that is those which vary on lengths greater than 100 km and on timescales greater than 1 sec. These include the formation of the magnetosphere itself, the pattern of plasma circulation within the magnetosphere, the existence and morphology of the various magnetospheric plasma populations, and the vast range of long-period magnetic pulsations recorded on Earth and in space.

Despite MHD's considerable success it should be noted that there are phenomena which it cannot explain and these often include much interesting physics. MHD breaks down where the magnetic field is very weak (witness equation 3), in particular in thin electric current sheets such as the magnetopause and neutral sheet. It may also become invalid where the plasma is very hot since the electron pressure gradient term (equation 1) may no longer be neglected, and its application becomes questionable in the presence of strong magnetic field-aligned currents $J_\parallel$ as the $\sigma^{-1}J_\parallel$ term (equation 16.1) may become significant, for example, above the auroral zone. In such cases it is necessary to use orbit theory and/or kinetic theory to gain a full understanding of the plasma behaviour.

16.2.2 General Structures

MHD explains well the large-scale features of the magnetosphere. In this section I first provide an introductory "road map" to the magnetosphere and then review in more detail those key plasma and magnetic field structures most easily explained by MHD.

16.2.2.1 Basic Characteristics

A vast comet-shaped magnetic cavity called the magnetosphere exists around the Earth because the solar wind (a tenuous (5cm^{-3} density) fast flowing (400 kms^{-1} speed) highly electrically conducting

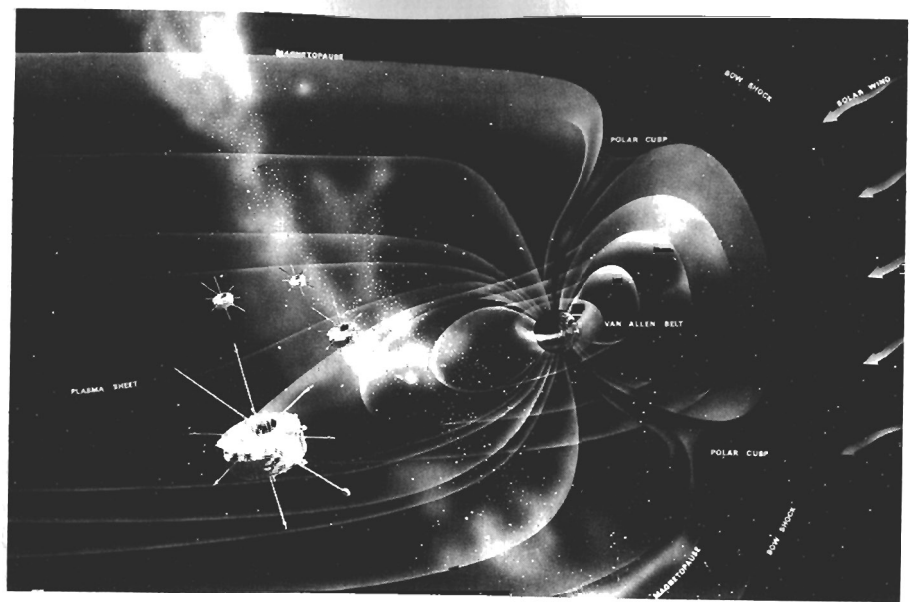


Figure 16.2. Important features of the Earth's magnetised space environment (magnetosphere) pictured with the four satellites comprising the European Space Agency's CLUSTER mission (from Wenzel et al., 1987).

plasma streaming radially away from the Sun) is largely deflected by the Earth's magnetic field. There is some momentum transfer, however, for the solar wind sweeps the Earth's field into a huge cylindrical magnetotail stretching for millions of kilometres into interplanetary space. Major features of the Earth's magnetosphere are illustrated in Figure 16.2.

The outer boundary of the magnetosphere, called the magnetopause, is a current sheet where the dynamic pressure of the solar wind (ρv^2, where ρ is mass density) balances the magnetic pressure ($B^2/2\mu$, where μ is the magnetic permeability of free space) of the Earth's dipole field. Its sunward point occurs on average about 10 Earth radii (R_E) upstream from our planet ($R_E = 6370$km is a commonly used distance measurement in magnetospheric research) but the boundary is very dynamic responding even to small changes in solar wind dynamic pressure. The Van Allen radiation belts of energetic ions and electrons (with energies reaching to a few 10^6eV for electrons and to a few 10^8eV for protons) form a torus surrounding the Earth at altitudes between 1 and $6R_E$.

In the centre of the magnetotail is the plasma sheet. Here disruptions can occur releasing huge (1.5×10^{11}W) quantities of magnetic

energy which generate bright auroras in the Earth's two polar ionospheres and propel giant ($75R_E \times 20R_E$) bubbles of hot plasma, called plasmoids, which are threaded and held together by closed loops of magnetic field, into interplanetary space. Two polar cusps occur at high ($76° - 80°$) magnetic latitudes above the Arctic and Antarctic. They are funnel-shaped regions separating geomagnetic field lines projecting to the dayside magnetopause from those which are swept back to form the magnetotail. They provide a splendid window for examining processes which transfer solar wind mass and momentum to the magnetosphere. A bow shock wave exists about $5R_E$ upstream of the magnetospheric cavity standing in the solar wind flow. This shock wave is analogous to that associated with supersonic aircraft and exists because the solar wind velocity exceeds the speed at which pressure waves can travel sunwards to deflect the solar wind around the magnetosphere.

A classification of magnetospheric structures in terms of either a magnetic field structure or a plasma structure is a simplifying division to make although, of course, the overall morphology usually depends on a combination of both types.

16.2.2.2 Magnetic Field and Electric Current Structure

The large-scale magnetic field configuration within the Earth's magnetosphere is now known in great detail as demonstrated by the recent global empirical model of Tsyganenko (1987). Field strengths in the magnetosphere range from 10^4 nT in the topside ionosphere down to 10 nT or less in the central plasma sheet. Magnetic fields are crucial for the existence and structure of many magnetospheric phenomena for they can affect plasmas in different ways: (a) they exert a force on plasma particles called the Lorentz force; (b) they can accelerate and heat a plasma through the release of stored magnetic energy; (c) they provide thermal insulation; (d) they produce anisotropies. The major magnetic field structures associated with the magnetosphere are the bow shock, magnetosheath, magnetopause and magnetotail. This classification is based upon the presence of electric currents $\mathbf{J}$ which, from Ampere's Law ($\nabla \times \mathbf{B} = \mu_0 \mathbf{J}$), produce characteristic magnetic field rotations across these regions; they are not necessarily the sites where magnetic pressure dominates the plasma pressure. Figure 16.3 shows a schematic cross-section through the noon-midnight meridian plane of the magnetosphere together with a cross-section through the magnetotail looking towards the Earth. The magnetic field topology for a southward directed

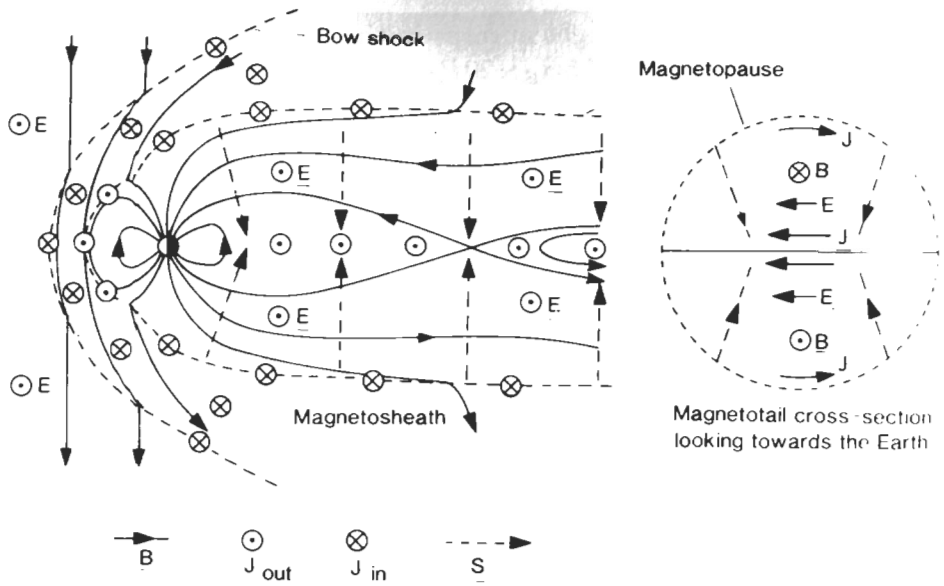


Figure 16.3. The major magnetic field and electric current structures of the magnetosphere when coupling with the solar wind is most intense (during southward directed interplanetary magnetic field conditions). The directions of the electric field and Poynting flux are also marked. A detailed description appears in the text (after Cowley, 1980).

interplanetary magnetic field (the most widely discussed and dynamically important direction) is shown together with the directions of **J**, electric field **E** and Poynting flux **S**.

Of these structures the magnetosheath and magnetotail have so far received little discussion. The magnetosheath is the region of compressed and heated solar wind plasma separating the bow shock from the magnetosphere, and remains one of the least explored regions of near-Earth space. The magnetotail is the nightside region where the Earth's dipole field is stretched antisunwards into a long cylindrical tail of diameter 40 R_E and length greater than 1000 R_E. It consists of two bundles of oppositely directed magnetic flux (called lobes) separated by a current sheet termed the neutral sheet.

Figure 16.3 shows the current is directed outwards on the dayside magnetopause and along the tail neutral sheet, and is directed inwards over the bow shock, tail magnetopause and in the magnetosheath. Since **E** is directed outwards everywhere in the left hand part of the Figure, $\mathbf{J} \cdot \mathbf{E}$ is > 0 on the dayside and in the tail so the electromagnetic field loses energy to the plasma (through reconnection). Elsewhere $\mathbf{J} \cdot \mathbf{E}$ is < 0 so the electromagnetic field gains energy at the expense of the plasma slowing. The net effect of this energy transfer is to extract energy from the solar wind, store it in the magnetotail field and then release it at the neutral sheet to produce the hot plasma which forms the plasma sheet.

Since $\nabla \cdot \mathbf{J} = 0$ in space plasmas, electric currents must flow in closed circuits. Current closure in the magnetotail is the simplest circuit to understand. Beyond a few tens of R_E down-tail the neutral sheet current closes in and around the tail magnetopause creating a "theta" current distribution (Figure 16.3 (right)). In the nightside region closer to the Earth, where the field is more dipolar, the neutral sheet current closes westward around the Earth giving rise to the ring current.

16.2.2.3 Plasma Sources and Populations

Magnetospheric plasmas differ widely in their temperature, density and chemical composition (see the Table). Temperatures vary from the relatively cool ($10^3 K$) plasmas of the topside ionosphere to the very hot ($5 \times 10^7 K$) plasma in the central plasma sheet, while densities range from the tenuous ($10^{-2} cm^{-3}$) tail lobe plasma to the dense ($10^4 cm^{-3}$) upper ionosphere. The dominant ionic species component is H^+ but minor amounts of He^+, He^{++}, O^+, O^{++}, D^+, N^+ are also observed together with heavier more exotic ions. With the exception of the topside ionosphere all magnetospheric plasmas are collisionless, that is the mean free path for collisions of their charged particles is greater than the characteristic linear scale of the plasma (see the Table). Thus the properties of magnetospheric plasmas are dominated by long-range electromagnetic forces rather than by the forces due to binary collisions. This is the reason why magnetospheric physics is such a difficult and challenging subject to study. A fine review of the various plasma populations within the magnetosphere is given by Cowley (1980).

There are two principal sources of magnetospheric plasma: the solar wind and the ionosphere (Figure 16.4). During intervals of quiet geomagnetic activity the solar wind is the dominant source, with plasma penetrating the magnetopause to produce, initially, boundary-layer plasmas covering the entire magnetospheric cavity surface. These plasmas consist mainly of ionised hydrogen,together with a few per cent of He^{++}, at energies between $100eV$ and $1keV$ and at a density of 10 cm^{-3}.

As geomagnetic activity rises, the ionosphere becomes increasingly

Table 1. *Basic parameters for magnetospheric plasmas.*

Plasma Parameter	Solar Wind at 1 A.U.	Low-Latitude Boundary Layer	Plasma Mantle	Tail Lobe	Plasma Sheet	Plasmasphere	Polar Wind (low-altitude)	Topside Ionosphere
Number density n_i (cm^{-3})	10	10	1	10^{-2}	1	10^3	0.1	10^4
Ion Temperature T_i (°K)	10^5	5×10^6	10^6	10^5	5×10^7	10^5	10^5	10^3
Plasma flow speed v (km s^{-1})	400	200	150	40	10	10	20	0.1
Main Chemical Composition	$H+$, $He{+}{+}$	$H+$, $He{+}{+}$	$H+$, $He{+}{+}$	$H+$, $O+$ $He+$, $N+$	$H+$, $He{+}{+}$ $O+$, $He+$, $N+$	$H+$, $O+$ $He+$, $N+$	$H+$, $O+$ $He+$, $N+$	$H+$, $O+$ $He+$, $N+$
Magnetic Field B (nT)	10	40	25	25	10	10^3	10^4	10^5
Scale length L (km)	10^8	10^4	10^5	10^5	10^4	10^4	10^4	10^3
Mean free path (km)	10^9	10^{12}	10^{12}	10^{10}	10^{15}	10^7	10^{11}	10^2
Thermal ion gyroradius (km)	10^2	10^2	10^2	10^2	10^3	1	0.1	10^{-3}
Plasma beta β	0.5	1	0.5	10^{-4}	10	10^{-3}	10^{-9}	10^{-8}

Figure 16.4. The two basic sources of magnetospheric plasma.

important as a source of magnetospheric plasma. Indeed during magnetic storm conditions it may contribute the dominant amount of mass in the inner magnetosphere (for example, see the AMPTE mission ion composition results of Gloeckler and Hamilton (1987)). There are two types of ionospheric plasma outflow. The first and more global source is called the polar wind (Figure 16.4) and occurs because gravity cannot contain the pressure of the Earth's ionosphere (it is analogous to the solar wind expansion from the Sun). The polar wind (see the Table) consists of low energy ($1 - 10eV$) protons and electrons flowing out along magnetic field lines at speeds of a few 10's kms^{-1} and at densities of order 0.1 to 1 cm^{-3}. The second type of outflow occurs above the Earth's two auroral zones (ovals surrounding the Earth at polar magnetic latitudes around 75°). This plasma outflow is accelerated to higher energies ($100eV$ to a few keV) corresponding to outflow speeds of a few $100kms^{-1}$ and contains both H^+ and O^+ with minor amounts of He^+ and N^+. The acceleration mechanisms responsible for this outflow are not easily explained by MHD and appear to be a combination of low-altitude transverse heating by ion cyclotron waves and upward acceleration by parallel electric fields. The solar wind and ionospheric plasma sources give rise to distinct plasma populations within the magnetosphere.

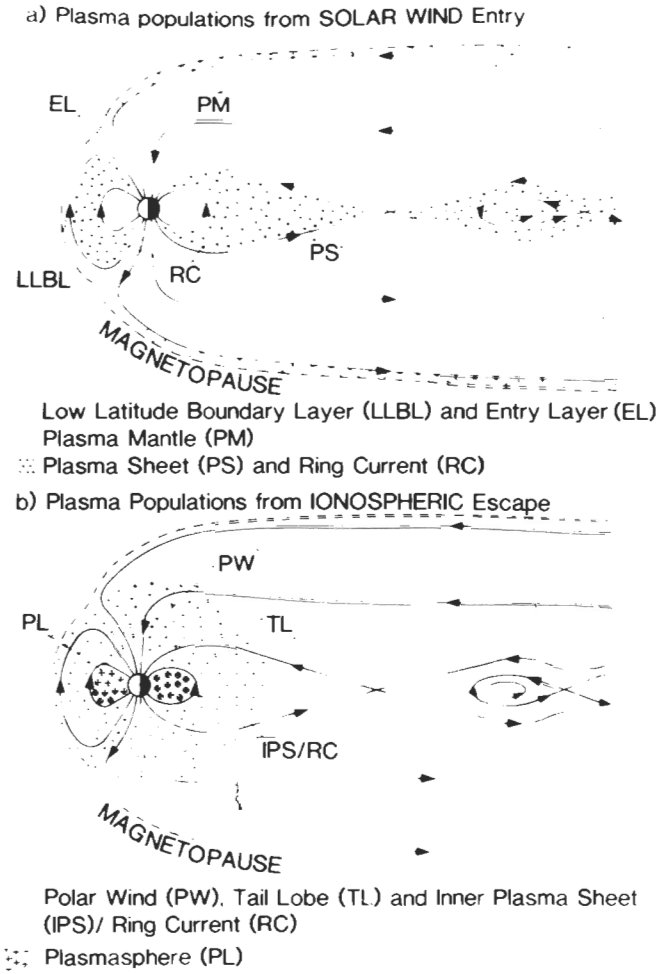


Figure 16.5. The primary plasma populations within the Earth's magnetosphere, distinguishing those arising from solar wind plasma entry from those coming from ionospheric plasma escape (based upon unpublished diagrams by S.W.H.Cowley).

These populations are illustrated in Figure 16.5 with the upper diagram (a) showing those populations established by the entry of solar wind plasma and the lower sketch (b) containing the magnetospheric plasma populations coming from ionospheric plasma escape.

Solar wind plasma (H^+ and He^{++}) contributes to three major plasma populations: (i) the low latitude boundary layer (LLBL) and entry layer (EL), (ii) the plasma mantle, and (iii) the plasma sheet and ring current. The former two come from direct solar wind access

across the magnetopause while the plasma sheet/ring current is set up by plasma mantle particles $\mathbf{E}\times\mathbf{B}$ drifting towards the magnetotail centre plane and then undergoing current sheet acceleration through reconnection. The basic characteristics of these plasmas are listed in the Table. The LLBL and EL are a few 1000 km wide and extend down to the ionosphere as the dayside cusp/cleft. These particles mirror and then flow out to form a boundary layer adjacent to the tail magnetopause called the plasma mantle which is a few R_E thick. The plasma mantle thickens downtail due to inward plasma transport, reaching the tail centre plane at a downtail distance beyond $100R_E$. The hot plasma sheet population flows either Earthward or tailward depending upon position relative to the tail magnetic neutral line. The Earthward flowing plasma sheet population gives rise to the ring current population in the quasi-dipolar inner magnetotail. The inner edge of the ring current is called the plasmapause and marks the boundary of those magnetic field lines which corotate with the Earth.

Ionospheric plasma (principally H^+, O^+, He^+ and N^+) contributes to four key plasma populations: (i) the polar wind, (ii) the tail lobe plasma, (iii) the plasmasphere, and (iv) the inner plasma sheet and ring current. The plasma characteristics of these regions are listed in Table 16.1. The polar wind has been described already. Its outflow extends to the dayside magnetopause but only reaches a few tens of R_E downtail (see Figure 16.5) due to its low temperature and consequent low flow speed along $\mathbf{B}$. The tail lobe plasma is of polar wind origin, being plasma which has penetrated onto magnetotail open field lines and is $\mathbf{E}\times\mathbf{B}$ drifting towards the tail neutral sheet. The plasmasphere is a high-density cold plasma in dynamic equilibrium with the ionosphere and corotating with the Earth; its outer boundary is the plasmapause where the plasma density drops by two or three orders of magnitude. Ionospheric plasma from the polar wind also contributes to the near-Earth plasma sheet and ring current. In addition, accelerated ionspheric ions of energy up to a few keV are injected into the plasma sheet from above the auroral zones.

16.2.3 Magnetospheric Processes

This section briefly reviews those plasma and magnetic field processes (e.g. magnetic reconnection, field-aligned currents and collisionless shocks) fundamental to the gross (MHD) structure of the magnetosphere.

16.2.3.1 Magnetic Reconnection and Plasma Circulation

Of all the plasma processes in magnetospheric physics magnetic reconnection must rank as the most important. It is the principal means by which solar wind momentum couples to the magnetosphere, this coupling occurring through the release of magnetic stresses with about $5 \times 10^{11} W$ of power being transmitted. It is also the primary channel for plasma mass exchange between the solar wind and dayside magnetosphere. In the nightside magnetosphere magnetic reconnection is responsible for the violent disruptions, called substorms, which are triggered in the centre of the Earth's huge magnetotail. Each substorm releases about $10^{15} J$ of energy leading to brilliant auroras above the Earth and the catapulting of a plasmoid into interplanetary space. In short, magnetic reconnection drives most of the plasma transport within the magnetosphere and thus controls the macroscopic distribution of plasmas and magnetic fields therein. Readable summaries of the topic are given by Cowley (1985) and by Lockwood et al. (1990b).

Magnetic reconnection is a concept formulated in the 1950s by solar physicists attempting to explain solar flares. It releases magnetic energy into plasma jets and particle heating and its key features are illustrated in Figure 16.6. The process occurs when two plasma regions having oppositely directed magnetic fields are pushed together at a current sheet. Provided these plasmas have a finite resistance, the field lines may break and reconnect in new combinations, resulting in a magnetic null situated at the centre of the 'X' in the magnetic field. Magnetic energy is converted into plasma motion (labelled 'jets') when the reconnected field lines are catapulted away by the magnetic tension ($\mathbf{J} \times \mathbf{B}$) force acting at the hairpin bend in the field lines. Sophisticated reconnection theories now exist (e.g. Priest and Forbes, 1986; Priest and Lee, 1990) which explain the phenomena in terms of either an MHD (fluid) or an individual particle viewpoint, both descriptions being closely similar.

The pattern of plasma transport driven by reconnection is depicted in Figure 16.7 for the noon-midnight meridian plane. This is called the 'open' magnetosphere model and is characterised by polar geomagnetic field lines extending across the magnetopause into interplanetary space rather than having both feet connected to the Earth as in the 'closed' magnetosphere model. The direction of plasma transport is marked by the heavy arrows. It starts with reconnection occurring at the dayside neutral line. The open magnetic field lines thereby created are pulled over the poles by the solar wind flow

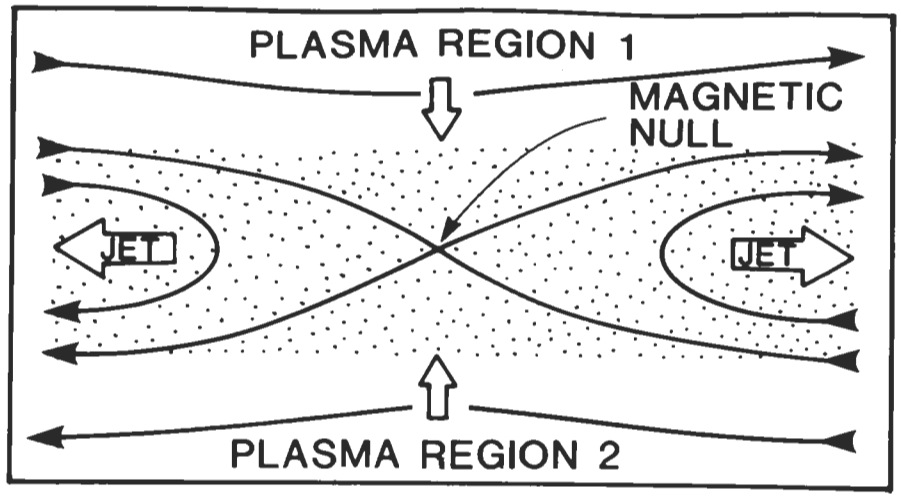


Figure 16.6. Basic physics of the magnetic reconnection process. Solid lines indicate the magnetic field direction while open arrows give the direction of plasma flow and the dotted region is a current sheet.

and stretched out into a long ($1000 R_E$) magnetic tail. These field lines sink ($\mathbf{E} \times \mathbf{B}$ drift) towards the tail centre plane where they reconnect again at the nightside neutral line leading to return flow towards the dayside on closed magnetic field lines. Thus reconnection drives a cyclic flow with open field lines being transported from the dayside to nightside and closed field lines being transported from the tail back to the dayside. The total cycle time is between 6 and 12 hours of which 2-4 hours is spent as an open field line with 4-8 hours as a closed one (Cowley, 1980). All magnetospheric magnetic field lines participate in this convective cycle, the only exception being those field lines within the plasmasphere. In practice the convection cycle is rarely steady and occurs usually in bursts.

16.2.3.2 Momentum Transfer and Field-Aligned Currents

Magnetic reconnection is the process by which solar wind mass and momentum enters the magnetosphere. But how does this momentum then couple through the magnetosphere ultimately to low altitudes where it produces plasma transport and heating in the polar ionosphere? Momentum transport is most easily understood by referring to the MHD momentum equation

$$\rho \frac{d\mathbf{v}}{dt} = \mathbf{J} \times \mathbf{B} - \nabla p + \rho \mathbf{g}, \qquad (16.4)$$

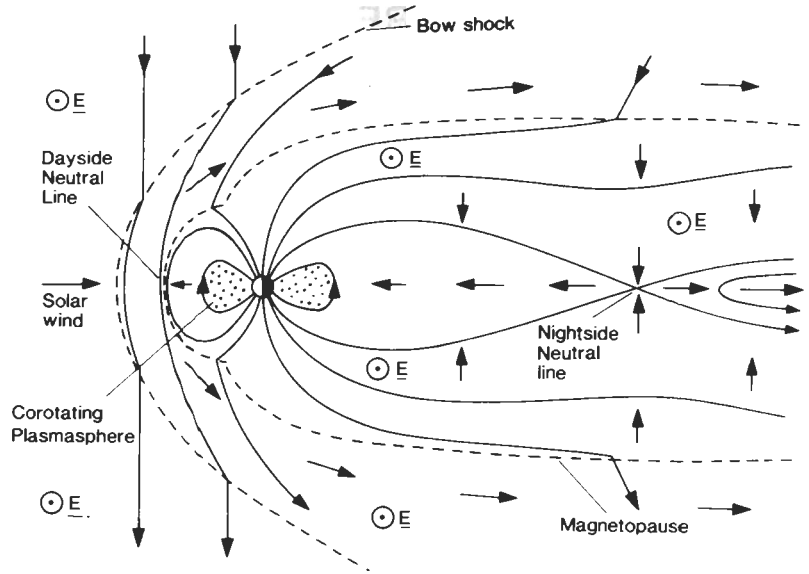


Figure 16.7. The global pattern of plasma transport within the magnetosphere driven by magnetic reconnection (based upon an unpublished diagram by S.W.H.Cowley).

where p is the plasma pressure and $\mathbf{g}$ is the acceleration due to gravity.

In the magnetosphere the gravitational force ($\rho\mathbf{g}$) term is negligible and the pressure gradient contribution (∇p) is often small so the electromagnetic Maxwell stress ($\mathbf{J}\times\mathbf{B}$) force term dominates in causing plasma motion $\mathbf{v}$. Substituting for $\mathbf{J}$ using Ampere's Law it follows that the Maxwell stress per unit volume is the sum of two terms:

$$\begin{aligned}\mathbf{J}\times\mathbf{B} &= \frac{1}{\mu_0}\Big(\nabla\times\mathbf{B}\Big)\times\mathbf{B}\\ &= \frac{1}{\mu_0}\Big(\mathbf{B}\cdot\nabla\Big)\mathbf{B}-\nabla\Big(\frac{B^2}{2\mu_0}\Big).\end{aligned}\qquad(16.5)$$

The first is the magnetic tension force per unit volume due to the change of $\mathbf{B}$ along $\mathbf{B}$; it is the force which tries to reduce any bend in the magnetic field. The right-hand term is the magnetic pressure force per unit volume. Thus the flux of perpendicular momentum transmitted by Maxwell stresses along $\mathbf{B}$ per unit area is given simply by $B_0\Delta B_\perp/\mu_0$ where B_o is the background magnetic field and $\Delta B_\perp$ is the local perpendicular magnetic field perturbation. Now in magnetospheric physics there is an alternative but equivalent description for momentum transfer along $\mathbf{B}$ involving electric currents rather than Maxwell stresses. In this alternative approach momentum coupling is achieved by magnetic field-aligned or Birkeland currents $\mathbf{J}_\parallel$. Figure 16.8 illustrates schematically how stress may be communicated by Birkeland current flow. It shows magnetic field lines convecting through the magnetosphere at the $\mathbf{E}\times\mathbf{B}/B^2$ drift speed due to their interaction with the solar wind. Ion-neutral collisions in the ionosphere cause the latter to be 'sticky' to this motion and a perpendicular field perturbation results above the ionosphere. The momentum input to the ionosphere per unit area is proportional only to $\Delta B_\perp$ and balances the $\mathbf{J}_p\times\mathbf{B}$ drag force which the Pedersen current flow ($\mathbf{J}_p$) exerts on the ionosphere. From Ampere's Law $\Delta B_\perp$ is proportional to $\mathbf{J}_\parallel$, so the momentum input along $\mathbf{B}$ is clearly also directly proportional to $\mathbf{J}_\parallel$. There is current closure between the Birkeland current, the magnetospheric driving current J_m and the ionospheric Pedersen current.

Electric coupling of the magnetosphere and ionosphere by field-aligned currents was suggested first by the Norwegian Birkeland in 1908 (Chapman and Bartels, 1940). Thanks to extensive observations by low-altitude polar-orbiting satellites from the mid 1970's, we now know that intense ($10^5 A$) large-scale Birkeland current sheets,

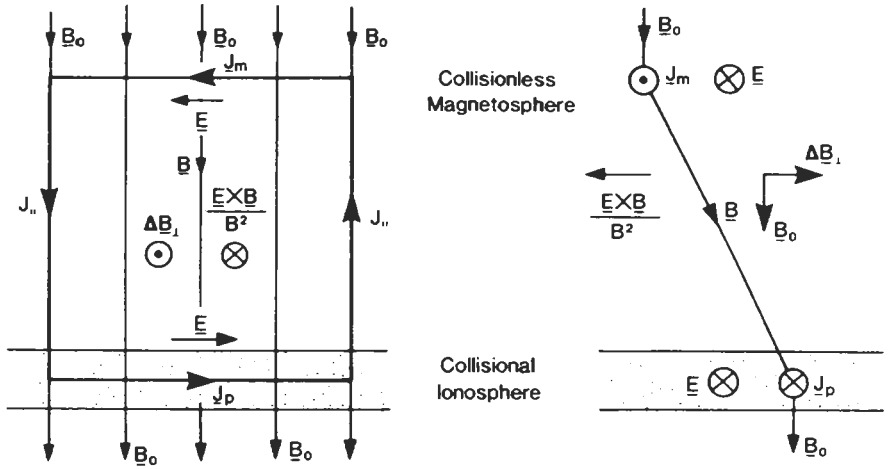


Figure 16.8. Diagram illustrating how momentum is transferred by field-aligned current flow (after Southwood and Hughes, 1983).

each of width about 200 km and carrying a current density of about $2\mu Am^{-2}$, do indeed electrically couple the high-latitude ionosphere with the outer magnetosphere. These currents arise from the circulation pattern of magnetospheric plasma described above since Birkeland current flow is colocated with plasma flow shear. This follows from simple MHD theory by taking the divergence of the frozen-in-field condition $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$ and using both Ohm's Law and the divergence-free condition of electric currents in space, to give for the ionospheric field-aligned current (Sato and Iijima, 1979)

$$J_{\parallel} = \Sigma_p \nabla_{\perp} \cdot \mathbf{E} = -\Sigma_p \mathbf{B} \cdot (\nabla \times \mathbf{v}), \qquad (16.6)$$

where Σ_p is the ionospheric height integrated Pedersen conductivity and the subscript '$\perp$' once again indicates the component perpendicular to the magnetic field. This equation for $J_{\parallel}$ is valid provided the ionospheric conductivity is uniform. It shows that Birkeland current sheets must be centred on regions of $\nabla \times \mathbf{v}$, with the direction of $\mathbf{J}$ determined by the sign of $\nabla \times \mathbf{v}$. There are five distinct Birkeland current regions: (1) the Region 1 and (2) the Region 2 systems which circle the pole at auroral latitudes; (3) the 'Cusp' region system; (4) the Harang midnight area of overlapping and multi-sheet Birkeland currents, and (5) the polar cap 'NBZ' system occurring when the interplanetary magnetic field is directed northward.

16.2.3.3 MHD Discontinuities and Waves

The study of MHD waves and discontinuities continues to attract interest. Although they are of marginal importance for the overall dynamics of the magnetosphere they do occupy a significant fraction of the theory of MHD plasma physics. Here I shall focus briefly on some key aspects of the Earth's bow shock and of hydromagnetic waves in general (see Siscoe (1983) and Southwood and Hughes (1983) for more thorough descriptions). Shock waves form one of four classes of MHD discontinuity (the others being contact discontinuity, tangential discontinuity and rotational discontinuity) across which mass, momentum and magnetic flux are conserved. There are three types of shock wave defined depending on whether the change in normal magnetic field B_n and tangential magnetic field B_t across the shock are zero or non-zero. These are parallel shocks ($B_t = 0$), perpendicular shocks ($B_n = 0$) and oblique shocks ($B_t \neq 0$, $B_n \neq 0$). Oblique shocks may be further divided into fast, slow and intermediate mode shocks. The Earth's bow shock is the most studied example of a shock wave in space with considerable advances in knowledge coming from the successful ISEE spacecraft mission (1977-87). Because it is a paraboloid of revolution it supports all three types of shock wave during steady interplanetary magnetic field conditions, as illustrated in Figure 16.9. In practice, however, the bow shock is usually defined as either quasi-perpendicular ($\theta_{B_n} \geq 45°$) or quasi-parallel ($\theta_{B_n} \leq 45°$) where θ_{B_n} is the angle between the upstream magnetic field direction and the shock normal.

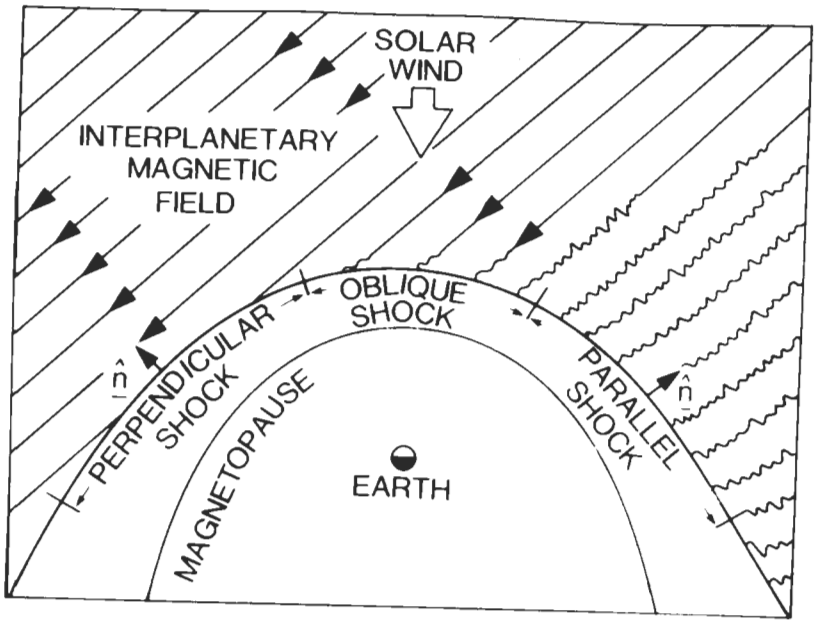


Figure 16.9. The Earth's bow shock illustrating how the character of the shock wave changes with longitudinal position under typical interplanetary magnetic field conditions.

The Earth's bow shock plays the role of converting a supermagnetosonic ($M_F > 1$) plasma flow (the solar wind) to a submagnetosonic ($M_F < 1$) plasma flow (the magnetosheath) where M_F ($= V/(V_A^2 + V_S^2)^{1/2}$) is the fast magnetosonic Mach number of the solar wind with V, V_A, and V_S being the ambient plasma flow, Alfvén and sound speeds, respectively. Thus the bow shock acts to slow the flow thereby increasing its density. It also dissipates some of the flow energy, converting it into heat. Particle acceleration occurs within the thin shock front with electrons and protons being accelerated to many times their original energy. Since charged particle collisions are negligible in producing the dissipation, the Earth's bow shock is described as a collisionless shock.

The Earth's magnetosphere supports an enormous range of hydro-

magnetic wave phenomena, with periods ranging from seconds to ten minutes. Such waves were first detected on the ground more than a century ago as geomagnetic pulsations (small fluctuations in the Earth's magnetic field). There are three distinct types of MHD wave called the transverse, fast and slow modes.

The transverse (also called the Alfvén , shear or intermediate) mode is the most significant hydromagnetic wave mode. Its group velocity is always parallel to **B** so it is field-guided and carries parallel current. Thus the transverse mode transmits stress along **B**, good examples being the coupling down to the ionosphere of stresses imposed by solar wind pressure variations impinging on the magnetopause (coupling which gives rise to surges in ionospheric plasma flow detectable with ground-based radars), and the coupling from the magnetotail associated with the sharp surge in plasma transport accompanying the initiation of a magnetic substorm (seen as a characteristic damped magnetic signal called a Pi2).

The behaviour and dynamic properties of the fast and slow mode waves are more subtle. The fast mode propagates isotropically and is compressional with magnetic and plasma pressures varying in phase. The slow mode is also compressional but particle and field pressures oscillate in antiphase. It cannot transmit energy across B and is similar to a sound wave in a pipe. The dynamic importance of the fast and slow modes is questionable but as pressure oscillations are observed within the magnetosphere, these modes clearly warrant attention.

16.3 Important Recent Developments

Research on magnetosphere MHD continues apace but recent progress has become focussed on certain topics. Many of these topics are of wider relevance in solar system plasma physics. Due to page constraints my choice of topics will perforce be selective and I apologise to readers who may feel that they or their favourite theme have been ignored or dismissed briefly. The controversies and the key remaining questions are highlighted and suggestions are made on how to proceed further.

16.3.1 Magnetic Reconnection

I shall divide this topic under three headings: dayside quasi-steady reconnection (QSR), flux transfer events (FTEs) and substorms.

16.3.1.1 Dayside QSR

Dayside QSR is inferred from sheets of accelerated plasma flow at the Earth's dayside magnetopause where the reconnection rate is constant for at least a minute. To date 36 magnetopause passes have been published (Sonnerup et al. 1981; Paschmann, 1984; Aggson et al. 1984; Paschmann et al. 1986; Gosling et al. 1986) where the plasma and magnetic field data satisfy the empirical "momentum stress balance test" for QSR. The best statistics on occurrence come from Paschmann et al. (1986) who, from a survey of 21 dayside magnetopause crossings by the AMPTE-IRM satellite, find that QSR occurs in about 50% of cases where a large ($> 90°$) magnetic shear exists across the magnetopause. These experimental findings leave little doubt that QSR is responsible for driving the bulk of magnetospheric motion: it supplies between 60% and 100% of the typical potential drop across the magnetosphere (Cowley, 1986). Good recent reviews on dayside QSR include those by Cowley (1986) and by Elphic (1987).

While little recent progress has been made observationally, significant theoretical advances have been made in connection with QSR. The most promising fresh development concerns the structure of the reconnection layer, the layer of MHD waves comprising the plasma outflow region emanating from the reconnection line (Heyn et al., 1988; Rijnbeek et al., 1989; Biernat et al., 1989, 1990; Shi and Lee, 1990). As illustrated in Figure 16.10a, these layers consist of a combination of Alfven waves A, slow shocks S and a contact discontinuity C. The diffusion region is indicated by the central dotted region and the subscripts 1 and 2 distinguish waves in the magnetosheath and magnetosphere parts of the layer.

The original Petschek (1964) reconnection theory has been extended to the more usual situation of a current sheet which separates magnetised flowing plasmas of different density and of arbitrary magnetic field direction. Rijnbeek et al. (1989) have used these theoretical results to compare the predictions of the Petschek model with satellite data from the first dayside magnetopause crossing identified as exhibiting QSR (8th. September 1978). Their findings are presented in Figure 16.10b with the left plot being the observed data and the right plot the model predictions. Parameters displayed are (from top to bottom): the plasma number density (cm^{-3}) the magnetic field strength B_t (nT) and angle α_B (degrees) in the tangential plane, and the corresponding plasma flow speed v_t (kms^{-1}) and angle α_V (degrees). The reconnection layer is the interval between the

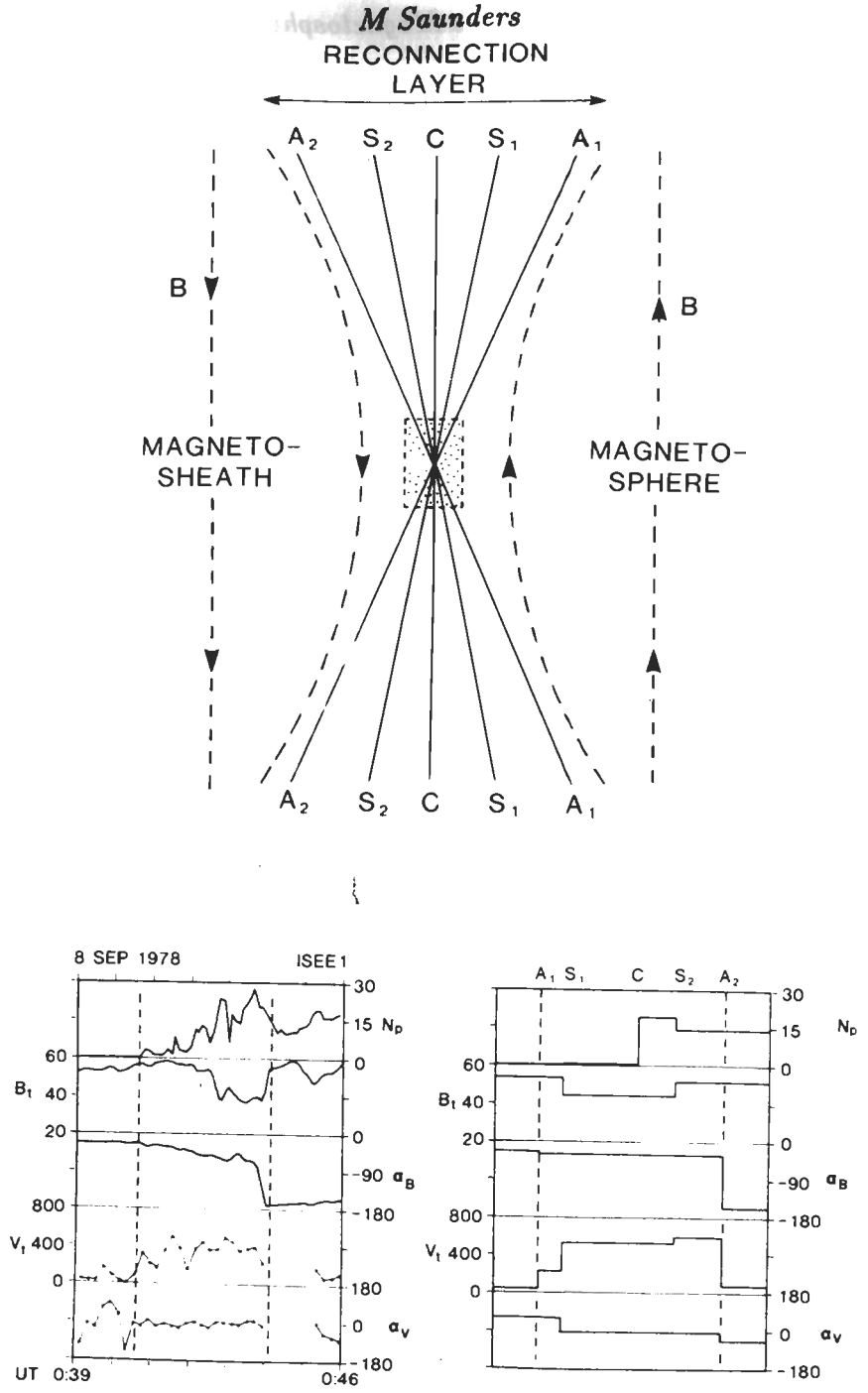


Figure 16.10. (a) Structure of the reconnection layer in the Petschek (1964) model (after Heyn et al. 1988). (b) Comparison of ISEE 1 plasma and magnetic field data (left) with the predicted structure of the reconnection layer (right) for a seven minute interval (after Rijnbeek et al. 1980).

dashed vertical lines. Clearly the observations and predictions agree sufficiently well to conclude that the Petschek model provides a useful framework for interpreting magnetopause data. This worthwhile work should be extended by refining the model and applying it to other data sets.

What are the important remaining questions for dayside QSR? Without doubt the most pressing outstanding issue concerns the other factor (in addition to large magnetic shear) for it to occur. One needs to investigate whether the magnetosheath plasma beta (Paschmann et al. 1986) or the difference in plasma densities and field strengths on the two sides of the magnetopause (Shi and Lee, 1990) may be controlling factors. Another important question requiring study is the identification of the diffusion region. This is vital for determining the source of the resistivity for reconnection to occur. As yet there is no definitive observation but identification should be possible with the high time resolution data available from AMPTE-UKS and IRM.

16.3.1.2 Flux Transfer Events

Reconnection does not always occur in a quasi-steady manner at the dayside magnetopause. In general, the rate of reconnection is variable and understanding the reason for its time-dependent nature is a major challenge in magnetospheric physics. Flux transfer events (FTEs) epitomise the time-dependent character of reconnection. FTEs were an unexpected discovery near the magnetopause in the early ISEE satellite data (Russell and Elphic, 1978), their identification largely being due to the introduction of a new coordinate system (boundary normal (LMN) coordinates) for plotting magnetopause magnetometer data. In this orthogonal system N is the estimated outward (positive away from the Earth) normal to the magnetopause with the latter lying in the LM plane such that L points roughly northward and M westward. FTE identification is based upon a characteristic bipolar signal in the magnetic field component B_N normal to the magnetopause. A northward moving event produces a positive/negative perturbation in B_N while southward motion creates a negative/positive signal. Four examples of FTE's are marked by the pairs of dashed vertical lines in the left-hand part of Figure 16.3. They illustrate the typical bipolar signal in B_N (in this case indicating southward motion) together with other key features such as an increase in field strength and the repitition in event occurrence (every 7 minutes).

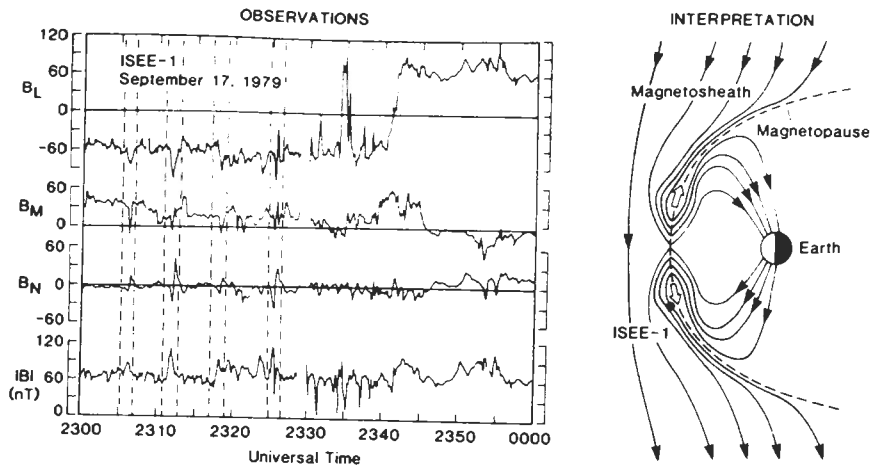


Figure 16.11. Flux transfer events: satellite observations from the magnetopause (left-taken from Berchem and Russell 1984) and the present popular interpretation (right-adapted from Scholer 1988).

Since their discovery over a decade ago, a variety of models have been proposed to explain their origin but the recent interpretational efforts by Scholer (1988) and by Southwood et al. (1988) appear to be the most promising, these being stimulated in part by the new observational findings from the AMPTE mission, notably from the multi-instrument study by Farrugia et al.(1988). This model stems from the suggestion by Saunders (1983) that FTEs might result from a pulse-like surge in the rate of reconnection. The right-hand part of Figure 16.11 illustrates the model diagrammatically based on a computer simulation (Scholer 1988). Reconnected magnetic field lines are drawn projected into the noon-midnight meridian plane. The bulges contain extra magnetic flux resulting from a surge in the reconnection rate. As these bulges move polewards in the direction of the open arrows due to the reconnection flow they will give rise to FTE signals in the overlying magnetosphere and magnetosheath field regions. For example, ISEE-1 would observe a negative/positive bipolar B_N signal similar to those recorded in the left-hand part of Figure 16.11. With one exception (the peak in field strength) all key characteristics of FTS's agree qualitatively with the model.

What are the key remaining questions concerning FTEs? Without doubt the major outstanding issue is the mechanism which triggers FTEs and/or causes surges in the rate of reconnection. In this regard, examining magnetosheath data for variations in either the plasma beta or dynamic pressure during FTE occurrence would be a worthwhile first study. An important related question is why are FTEs repetitive (every 8 minutes on average - Rijnbeek et al. 1984)? Answers to some of these questions may emerge soon but 3D spatial measurements from CLUSTER will certainly be needed before FTEs can be properly understood at the magnetopause. Hence the value of searching for the ionospheric signature of FTEs (within the polar cusp ionosphere) in forwarding our immediate understanding. The search for a conclusive identification has developed into one of the most exciting and controversial topics in magnetospheric physics. The most convincing evidence to date for such signatures comes from recent studies of the 'midday-auroral breakup' phenomenon (Lockwood *et al*, 1989; Sandholt *et al*, 1990b). Transient auroral forms are observed having many of the features expected of FTE's: these include occurrence predominantly during southward IMF conditions; east-west motion controlled by the IMF B_y component; 3-15 minute recurrence time; spatial scales of 50 - 500 km in latitude by 500 - 2000 km in longitude. Futhermore, individual features are associated with large electric potentials (25 - 55 kV) indicating that momentum coupling at the magnetopause is taking place in large bursts. Further studies of the dayside aurora should be productive but definitive identification of the FTE ionospheric signature must surely await simultaneous magnetopause and ground-based observations of the phenomena.

16.3.1.3 Substorms

The violent disruptions, called substorms, which occur a few times each day in the centre of the Earth's plasma sheet, must rank as the most spectacular phenomenon in magnetospheric physics. A substorm releases vast (10^{15}J) quantities of magnetic energy stored within the Earth's magnetotail, this energy being accumulated as a result of stresses imposed on the magnetosphere by the solar wind. This energy release produces a variety of phenomena, including dramatic auroral breakups, the injection and enhancement of ring current energetic particles, the catapulting of a plasmoid into interplanetary space and global ground magnetic disturbances.

Although substorm effects are well documented, with the significance of dayside reconnection as an instigating factor being widely

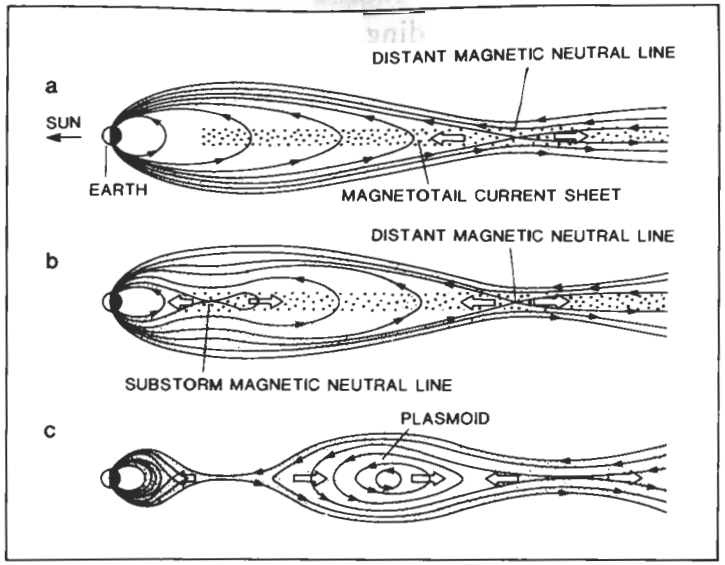


Figure 16.12. Sketches of the central portion of the Earth's magnetotail in the noon-midnight meridian plane showing the chain of dramatic disruptions, culminating in the ejection of a plasmoid, which accompany a substorm in the neutral-line model. Solid lines are magnetic field lines and open arrows indicate the direction of plasma flow (after Hones, 1984).

agreed, detailed models for their occurrence and growth remain a topic of much contention. There are two principal classes of model: the neutral-line model (see, for example, Hones (1984)), and the boundary layer dynamics model (Rostoker and Eastman, 1987) of which the thermal catastrophe model (Smith et al., 1986; Goertz and Smith, 1989) is a special case. The neutral-line model is the oldest and most widely accepted substorm model and its main features are sketched in Figure 16.12. Initially there is a distant magnetic neutral line situated about 100 R_E downstream from the Earth. At the start of the substorm a near-Earth or 'substorm' magnetic neutral line forms about 15 R_E downstream from the Earth, leading to the rapid disconnection of a plasmoid which is propelled downstream by a magnetic slingshot force at speeds of around 500 kms^{-1}. Although the model is not expressed in an analytical quantitative form, it has received extensive support from the recent MHD computer simulations of Walker and Ogino (1989). Its weaknesses generally involve phenomena seen at the feet of field lines in the ionosphere.

An interesting new development related to the substorm neutral line model has been proposed by Richard et al. (1989). They present two-dimensional MHD computer simulations highlighting the relevance of the magnetic island coalescence instability (Finn and Kaw, 1977) for magnetotail reconnection. In particular, their simulations suggest that plasmoid formation can occur through the coalescence of several magnetic islands rather than directly through the large-scale reconnection at a neutral line implicit in Figure 16.12. As plasmoids are three-dimensional structures, further work is necessary to extend the Richard et al. (1989) simulations to three dimensions. It is also necessary to demonstrate that coalescence occurs sufficiently fast to match the 5-10 minute timescale for the expansion phase of a substorm.

What are the major unanswered questions in substorm research? It seems that the number of theories is continuing to grow without a consensus being reached. To this reviewer there appears little doubt that substorm initiation is closely related to magnetotail reconnection which is described best by the substorm neutral line model. It is difficult, however, to see the debate being properly resolved prior to the availability of extensive global spaceborne and ground-based observations.

16.3.2 Dayside Birkeland Currents

With the polar ionosphere around noon now becoming a major focus for international space plasma research, interest is growing on the nature of the Birkeland current sheets which couple solar wind momentum to this region. I shall describe progress on this topic under two headings: the origin and morphology of dayside Birkeland currents and the relevance of Birkeland currents for driving dayside polar ionospheric motion.

16.3.2.1 Origin and Morphology

Three major regions of Birkeland current are reported above the dayside polar ionosphere during southward IMF conditions: the Region 1 and Region 2 systems which practically encircle the magnetic pole, and the cusp Birkeland current system which, by tradition, is classified as a crescent-shaped area found poleward of Regions 1 and 2. In the northern polar ionosphere the Region 1 current flows down on the dawnside of noon, while the Region 2 system lies at lower magnetic latitudes and is oppositely directed to the adjacent Region 1. The intensities of these systems differ markedly, however, with the dayside Region 1 and cusp Birkeland currents being far superior

to the Region 2 system. The morphology and origin of these two dominant systems has seen significant progress recently, as I now describe.

Self-consistency between Birkeland current flow and ionospheric plasma flow is essential (see equation 16.6). New models for the plasma flow in the ionospheric polar regions have recently become available (Heppner and Maynard, 1987) which far exceed the accuracy of previous models. These models are empirically-based incorporating extensive recordings from the Dynamics Explorer 2 satellite which made polar crossings in all magnetic local time zones. Figure 16.13 displays the plasma flow pattern in the northern polar ionosphere during conditions of negative $IMF\ B_Y$ and moderate geomagnetic activity. Superimposed are the associated Birkeland current sheets estimated using equation (6); these sheets are colocated with the regions of plasma flow shear ($\nabla \times \mathbf{v}$). Analytical calculations also reveal a similar Birkeland current pattern to Figure 16.13 (Rich and Maynard, 1989).

The significant new feature of Figure 16.13 is that the poleward Region 1 current in the noon sector between 0930 and 1430 magnetic local time has the same polarity and characteristics as the 'traditional' cusp Birkeland current for negative $IMF\ B_Y$ conditions. The key difference, of course, is that the cusp Birkeland current system is no longer a spatially separate region as traditionally pictured (see, for example, Potemra and Saflekos, 1979) but is the extension in longitude of the Region 1 Birkeland current from dawn. A similar result holds for positive $IMF\ B_Y$ conditions in the northern hemisphere (not shown) for examining the 'improved' plasma flow model (Heppner and Maynard, 1987) for regions of velocity shear again shows clearly that the poleward Birkeland current sheet near noon again comes from the projection in longitude of the Region 1 current system - though this time it is from dusk rather than from dawn. Thus a more accurate name for the traditional 'cusp' Birkeland current would be the 'poleward Region 1 B_y' Birkeland current.

The morphology for the dayside Birkeland currents proposed in Figure 16.13 fits easily with the recent theoretical model for their origin (Saunders, 1989a) which argues that these currents arise from the magnetic field bending accompanying the motion of reconnected open magnetic field lines as they shorten and convect over the dayside magnetopause (by 'shortening' I mean the straightening which comes from the release of field tension at the sharp hairpin bend where a reconnected field line crosses the magnetopause). The motion of a

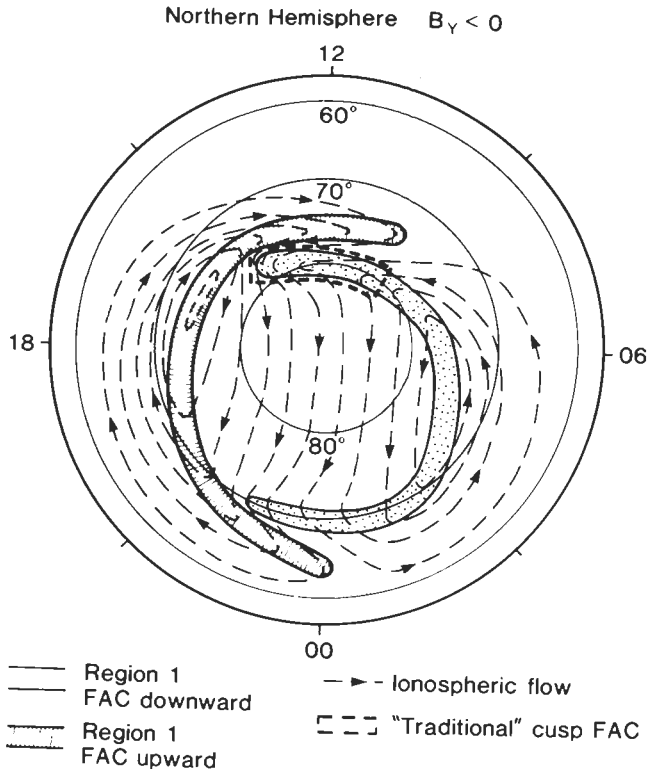


Figure 16.13. Plot of magnetic latitude versus magnetic local time for the northern polar ionosphere showing the recently published empirical plasma transport model for negative $IMF\ B_Y$ conditions (Heppner and Maynard, 1987). The Birkeland currents which, perforce, should accompany this ionospheric plasma flow pattern are shown superimposed; witness that the 'traditional' cusp Birkeland current is the extension towards dusk of the dawnside Region 1 system.

single reconnected dayside field line is illustrated in Figure 16.14. Initially this field line has a large east-west component to its motion due to the dominant effect of field tension. Later, as the field line straightens, the magnetosheath flow gains in importance, and the field line is pulled antisunwards into the polar cap. Each time the field line changes its direction of motion, the magnetic field must tilt and this tilting gives rise to the two dayside Region 1 Birkeland current systems shown in Figure 16.13.

16.3.2.2 Driving Dayside Polar Ionospheric Motion

Plasma transport in the dayside polar ionosphere is driven by the coupling of stresses imposed by the solar wind at the magnetopause. Until recently it was generally believed that this momentum input

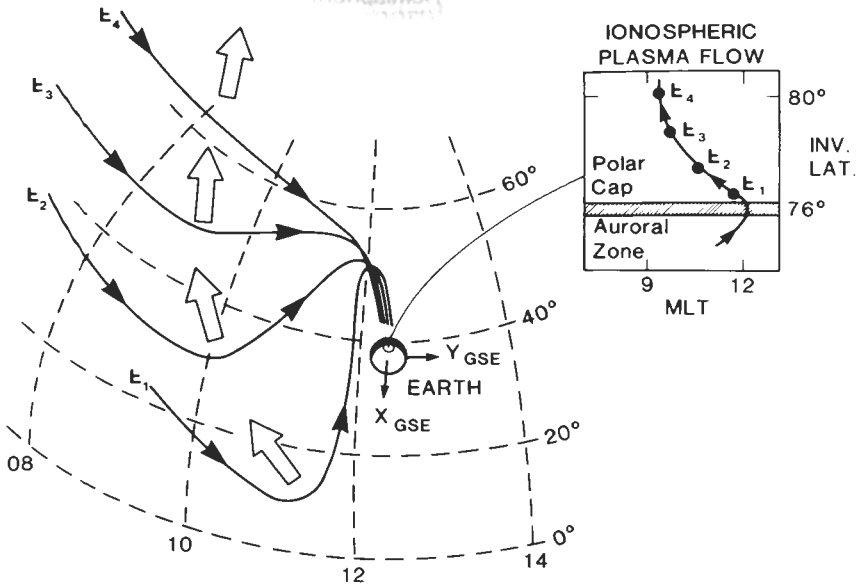


Figure 16.14. View of the northern dayside magnetopause with lines of solar magnetic latitude and local time marked in 200 and 2-hour intervals, respectively. The Figure is drawn for positive $IMF\ B_Y$ with the large arrows showing the direction of motion of a shortening reconnected field line at four roughly equispaced times (labelled t1 to t4). The change in direction of this motion (see text for explanation) maps to the ionosphere (see insert) and is the site of the two oppositely-directed Region 1 Birkeland currents near noon. The equatorward (downward-directed) Region 1 system projects to the polar cap boundary, while the poleward (upward-directed) system maps to where the field line completes its shortening (from Saunders, 1989a).

was spread across the entire dayside polar cap (see, for example, Vasyliunas (1975)). However, new observations by the EISCAT radar in northern Scandinavia (Lockwood and Cowley, 1988; Lockwood et al. 1990a) have brought this view into question. Their exciting discovery links the response time of plasma flows in the dayside polar ionosphere to IMF changes impinging on the magnetopause. In particular, following northward turnings of the IMF these plasma flows disappear within 5-10 minutes. This rapid decay implies that only newly-reconnected field lines impart significant momentum to the dayside polar ionosphere. Lockwood and coworkers note that the region of prime momentum input to the dayside polar ionosphere is restricted to a thin crescent shaped area centred on noon which corresponds roughly in size to where the two adjacent Region 1 Birkeland current sheets are observed (Figure 16.13).

The reason why the momentum input for driving dayside ionospheric motion is so localised is described by Saunders (1989b). From equation (16.5), momentum input to the ionosphere is significant only where $\Delta B_\perp$, the local perpendicular magnetic field perturbation is substantial. All the low-altitude satellite data published to date (see, for example, Bythrow et al. 1988) show that the dominant $\Delta B_\perp$ near noon is colocated with the two Region 1 Birkeland current systems described above. Furthermore it takes 5-10 minutes for a reconnected field line to shorten and move across this region (Saunders, 1989a), a timescale in good agreement with the duration of momentum input found experimentally by Lockwood and coworkers. Thus the momentum input for exciting plasma transport in the dayside high-latitude ionosphere comes from reconnected field lines shortening over the dayside magnetopause.

16.3.3 Shocks and Hydromagnetic Waves

As in Section 16.2.3.3 I shall discuss both these topics under a common Section heading, in each case focussing, for brevity, on just one or two of the major problems of recent interest. In the case of shocks it will be the relevance of Fermi acceleration for particle acceleration around the bow shock front, while for hydromagnetic waves I shall focus on recent advances concerning MHD waves near the magnetopause and in the magnetosheath. For broader reviews readers are referred (for shocks) to Schwartz (1985), Formisano (1985) and to the two AGU monographs edited by Stone and Tsurutani (1985) and by Tsurutani and Stone (1985), and (for hydromagnetic wave research) to Southwood and Hughes (1983) and Pilipenko (1990).

16.3.3.1 Fermi Acceleration

The question of particle acceleration through Fermi acceleration around shock fronts is probably the best example of an MHD problem where plasma physicists, space physicists and astrophysicists are interacting successfully. Its study extends from the astrophysical study of cosmic ray acceleration to the plasma physics theory of how strong MHD wave turbulence resonantly interacts with energetic ions. At the Earth's bow shock it is the so-called 'diffuse ions', the nearly isotropic ions of energy per charge between 30 and 200 keV observed in the quasi-parallel (Section 2.3.3) shock region, which are believed to come from Fermi acceleration. The principle of Fermi acceleration (Fermi, 1954) is that particles trapped between two magnetic mirrors will be accelerated if the distance between

these mirrors decreases sufficiently slowly to satisfy

$$\frac{\tau_B}{B}\frac{\partial B}{\partial t} \ll 1, \qquad (16.7)$$

where τ_B is the bounce period and $\partial B/\partial t$ is the rate of change of the field seen by the particle. The basic physics of Fermi acceleration applied to the Earth's bow shock is sketched in Figure 16.15 and involves ions reflecting back and forth across the shock from large-amplitude MHD wave turbulence. These waves, which occur only in the quasi-parallel upstream region, are believed to reflect the ions through pitch-angle scattering processes. In the shock frame of reference, upstream waves are swept towards the shock at super-Alfvenic speeds and transmitted through it emerging into a sub-Alfvenic flow where they propagate more slowly. Due to the upstream and downstream wave distributions being in relative motion, an ion will steadily increase its energy as it bounces back and forth across the shock from scattering off these waves. The total energy gain depends on the number of times it crosses the shock so Fermi acceleration is more efficient the larger and the longer lived the shock system (see Figure 16.15).

The Earth's bow shock, however, is not an easy environment in which to test the theory of Fermi acceleration, partly because the energetic particles seen nearby may have come from other acceleration processes inside the magnetosphere. Thus to compare theory with observations it is preferable to use an interplanetary shock. In addition to being 'cleaner' energetically, interplanetary shocks have the added benefits of being much larger than the Earth's bow shock and of having θ_{B_n} constant over a wider area of the shock front; effects which maximise the Fermi acceleration energy gain. A detailed empirical study has recently been made by Kennel and coworkers with results published in a series of papers culminating in Kennel et al. (1986) (see also Kennel (1987) and Terasawa and Scholer (1989)). This shock was recorded by the ISEE-3 spacecraft 220 R_E sunward of the Earth's bow shock. In what proved a difficult study Kennel et al. found close agreement between the predicted and observed values of the total energy density of the shock-accelerated ions (energies below 300 keV) assuming Fermi acceleration to be the source.

Despite the notable results of Kennel and co-workers, at least three major problems remain. The first is the matter of source particles for the Fermi acceleration process; in particular, what determines which particles are chosen from the bulk thermal population to undergo

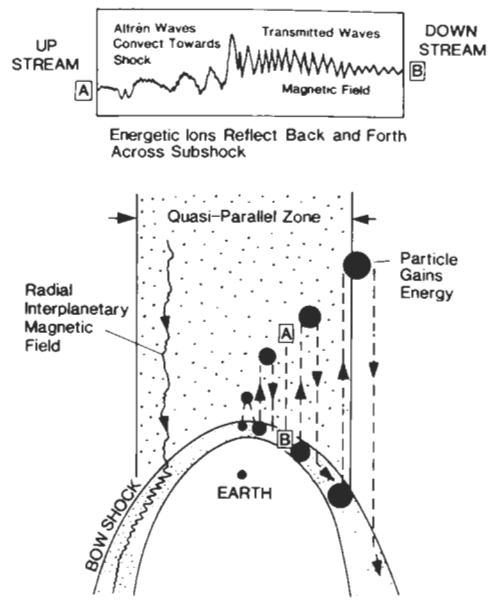


Figure 16.15. Schematic of how Fermi acceleration may operate at the Earth's bow shock. The top panel shows compressional Alfven waves propagating towards and through the shock, while the lower diagram illustrates the path and energy gain of a particle undergoing Fermi acceleration for the special case of a radial interplanetary magnetic field. The gain in energy is indicated by the growth in size of the solid circles (adapted from Kennel, 1987).

Fermi acceleration? Answering this question requires a clearer picture of the dissipation processes occurring in the shock front. The second question is the 'reflection problem'. We know that strong ($\delta B/B \sim 1$, where δB is the wave amplitude) MHD wave turbulence is necessary to give the required pitch angle diffusion for scattering and reflection to occur, but the precise details of the physics involved remains poorly understood. One possibility is that particle mirroring takes place from local maxima in the turbulent magnetic field. The third and most difficult issue to solve theoretically is how some particles at shock fronts are accelerated to energies much higher than predicted by simple (quasi-linear) Fermi acceleration theory. It appears that the scattering rate from the strong MHD turburlence is

particle energy dependent. This and other non-linear features of the Fermi acceleration process leave much scope for future research.

16.3.3.2 MHD Waves

The importance of MHD waves for particle acceleration near shocks has been described above. More recent advances in MHD wave research have come from near the magnetopause and in the magnetosheath, regions which are crucial for a complete understanding of solar wind-magnetosphere coupling. It has long been known that Kelvin-Helmholtz (K-H) waves can provide viscous momentum transfer across the magnetopause of up to 10 kV but the question of how common such waves are at the magnetopause has long been uncertain. Now Pu Song et al. (1988) in a statistical study of 12s averaged magnetometer records for the entire (10 year) ISEE-satellite magnetopause data base, suggest that the K-H instability plays only a monor role in causing long-period (> 2 minute) surface waves on the magnetopause. They argue that the majority of long period waves on the magnetopause come from variations in solar wind dynamic pressure and/or from reconnection.

A second area of recent observational advance related to the magnetosphere boundary concerns the report of an MHD slow mode wave, about 0.4 R_E thick, just exterior to the magnetopause. Pu Song et al. (1990) observe evidence for this wave (antiphase variations in magnetic field strength and plasma density) on more than half of the 26 ISEE subsolar satellite passes which they examined. Further study of this wave mode should be undertaken as it could be important for the nature of the magnetosheath plasma flow near the magnetopause.

The magnetosheath has, until recently, been one of the least explored regions of the magnetosphere. This is regrettable since much interesting science occurs in the magnetosheath. Now initial studies of its disturbed magnetic state are being undertaken to determine the origin of its magnetic activity by first characterising its dominant MHD wave modes and their direction of propagation. For example, Gleaves and Southwood (1990), using dual satellite ISEE magnetic field and plasma data identify three types of MHD disturbance: Alfven and slow mode waves propagating across the flow in the inner magnetosheath, and Alfven waves propagating downstream and along the field in the outer magnetosheath. Based on this limited study they conclude that the major source of magnetosheath wave activity is localised MHD disturbances on the bow shock.

16.4 Looking Ahead

I hope this chapter has demonstrated the vitality of recent research in our study of magnetosphere MHD. Several worthwhile advances have come in recent years, including key developments (both observational and theoretical) in our understanding of flux transfer events, clarification on the structure of reconnection layers and on the origin and morphology of high-latitude Birkeland current systems, and the recognition that Fermi acceleration probably occurs at the bow shock.

Looking ahead, an exciting future indeed lies in prospect with a variety of new satellite missions, novel experiments and research initiatives either planned or underway. The area offering the best prospects for substantial progress in the immediate few years is the polar cusp ionosphere, in particular its use as a window for monitoring solar wind-magnetosphere coupling processes at the magnetopause. Until 3D measurements from CLUSTER become available, our best hope for a deepened understanding of complex magnetopause phenomena such as flux transfer events is through observing their ionospheric signature. The polar cusp ionosphere also provides the opportunity for determining the role of ionospheric conductivity in affecting the initiation of dayside polar ionosphere plasma transport. In addition, examining statistics on the presence of the 'traditional' cusp Birkeland currents (renamed here the Region 1 B_y Birkeland currents) would provide a quantitative check on the occurrence of quasi-steady reconnection at the dayside magnetopause.

New observations of the polar cusp ionosphere will come from the PACE and Spitzbergen polar cap radars and GEM. PACE (Polar Anglo-American Conjugate Experiment) consists of two coherent radars located at the conjugate sites of Halley Bay in Antarctica and Goose Bay in Labrador. These radars can measure horizontal plasma flows in the F-region ionosphere over the entire polar cusp with 90s time resolution. The Spitzbergen polar cap radar is a sophisticated international experiment recommended for construction by 1995 to monitor the northern polar ionosphere at magnetic latitudes between 78° and 83°. It will extend the viewing area of the successful EISCAT radar polewards and also provide the high-resolution observations necessary for studying the dynamic coupling processes occurring at the magnetopause. GEM (Geospace Environment Modelling) is a major new research programme being implemented between 1990 and 1995, a key component of which will

be the use of ground and low altitude observations for monitoring phenomena at the magnetopause.

Perhaps the most exciting future development of all is the move away from single-point measurements of the magnetosphere towards multi-satellite and global observations. Such observations will provide an invaluable perspective on the time and spatial behaviour of macroscopic magnetosphere processes and, indeed, of the entire magnetosphere (Williams, 1990). Examples of such future multi-satellite missions are the Global Geospace Science (GGS) programme, part of the International Solar Terrestrial Physics (ISTP) programme, and SOHO and CLUSTER, the first cornerstone mission of ESA's long-term space plan, the so-called 'Horizon 2000' project. GGS consists of four satellites located in key regions (upstream solar wind, the polar cusp, the near equatorial plane and the magnetotail) of the magnetosphere to study its energy flow. The CLUSTER mission comprises four identical space probes orbiting at various separations through the Earth's magnetosphere. It will permit the first ever three-dimensional point measurements of magnetised space plasmas, thereby allowing important quantities such as $\nabla \times \mathbf{B}$, $\nabla \times \mathbf{v}$, $\nabla \cdot \mathbf{B}$ and $\nabla \cdot \mathbf{S}$ (giving information on electric currents, plasma vorticity, calibration of magnetometers and energy flow, respectively) to be measured directly for the first time. Both these multi-satellite missions will be supported by extensive ground-based observing programmes.

Among the experimental techniques already supplying global data are cameras giving global images of the aurora at optical and UV wavelengths (flown on the DMSP, HILAT, Polar Bear, DE-1 and Viking satellites), and detailed measurements of energetic neutral atoms (ENA) which provide global images of the hot magnetospheric plasma population (Roelof et al., 1985; Roelof, 1987). Using ENA measurements from instruments flown on the IMP and ISEE satellites, Roelof and coworkers have obtained remarkable global images of the ring current including pictures of its decay. Clearly both these techniques offer considerable future promise, especially when higher time resolution measurements become available allowing dynamic features in both the aurora and ring current to be studied.

Two final topics where significant developments are expected soon are computer simulations and the development of common ground with solar MHD. Although data analysis and theory still represent the major research techniques (as emphasised in this chapter), the role of MHD simulations in modelling plasma processes in three dimensions and at various time resolutions is growing steadily. The ability to 'fly' satellites through an MHD computer code and reproduce an observed data set must surely represent the ultimate test for whether an MHD process is completely understood. The magnetised space plasma environments of the Earth and the Sun have, by tradition, been studied separately. However, there is a 'magnetosphere-solar connection' in that the two environments share many common physical phenomena, including magnetic reconnection, filamentary structures, shocks, hydromagnetic waves and 'wind outflow'. Without doubt, scientific advances will come from Earth and Solar plasma physicists working together and pooling their knowledge and understanding of these processes in the future ESA CLUSTER/SOHO programme (to be launched in 1995). These are the future prospects in a fascinating and rapidly progressing subject where important advances are envisaged during the 1990s.

Acknowledgements.

This chapter was written during the course of a Royal Society 1983 Research Fellowship. Much of the material in Section 16.2 has been developed through close association with colleagues at Imperial College, notably Stan Cowley and David Southwood. I am grateful to Neal Powell for help in preparing some of the Figures.

References

Aggson, T.L., Maynard, N.C., Ogilvie, K.W., Scudder, J.D. & Gambardella, P.J. (1984) *Geophys. Res. Letts.*, **11**, 8-11.

Berchem, J. & Russell, C.T. (1984) *J.Geophys. Res.*, **89**, 6689-6703.

Biernat, H.K., Heyn, M.F., Rijnbeek, R.P., Semenov, V.S. & Farrugia, C.J. (1989) *J. Geophys. Res.*, **94**, 287-298.

Biernat, H.K., Heyn, M.F., Rijnbeek, R.P., Semenov, V.S. & Farrugia, C.J. (1990) *Annales Geophysicae*, **8**, (1), 69-78.

Bythrow, P.F., Potemra, T.A., Erlandson, R.E., Zanetti, L.J. & Klumpar, D.M. (1988) *J.Geophys. Res.*, **93**, 9791-9803.

Chapman, S. & Ferraro, V.C.A. (1931) *J. Geophys. Res.*, **36**, 77-97, 171-186.

Chapman, S. & Ferraro, V.C.A. (1932) *J. Geophys. Res.*, **37**, 147-156, 421-429.

Chapman, S. & Ferraro, V.C.A. (1933) *J.Geophys. Res.*, **38**, 79-96.

Chapman, S. & Bartels, J. (1940) In *Geomagnetism.* Oxford: Oxford University Press.

Cowley, S.W.H. (1980) *Space Sci. Rev.*, **26**, 217-276.

Cowley, S.W.H. (1985) In *Solar System Magnetic Fields*, (edited by E.R.Priest), Dordrecht: D.Reidel. pgs.121-155.

Cowley, S.W.H. (1986) *J. Geomag. Geoelect.*, **38**, 1223-1256.

Elphic, R.C. (1987) *Rev. Geophys.*, **25**, 510-522.
Fairfield, D.H. (1990) *J.Atmos.Terr.Phys.*, in press.
Farrugia, C.J., Rijnbeek, R.P., Saunders, M.A., Southwood, D.J., Rodgers, D.J., Smith, M.F., Chaloner, C.P., Hall, D.S., Christiansen, P.J. & Wooliscroft, L.J.C. (1988) *J. Geophys. Res.*, **93**, 14465-14478.
Fermi, E. (1954) *Astrophys. J.*, **119**, 1-6.
Finn, J.M. & Kaw, P.K. (1977) *Phys. Fluids*, **20**, 72-78.
Formisano, V. (1985) In *Future Missions in Solar, Heliospheric and Space Plasma Physics*, ESA SP-235, 83-100.
Gleaves, D.G. & Southwood, D.J. (1990) *J.Geophys. Res.*, in press.
Gloeckler, G. & Hamilton, D.C. (1987) *Physica Scripta*, **Vol. T18**, 73-84.
Goertz, C.K. & Smith, R.A. (1989) *J.Geophys. Res.*, **94**, 6581-6596.
Gold, T. (1959) *J. Geophys. Res.*, **64**, 1219-1224.
Gosling, J.T., Thomsen, M.F., Bame, S.J. & Russell, C.T. (1986) *J. Geophys. Res.*, **91**, 3029-3041.
Heppner, J.P. & Maynard, N.C. (1987) *J.Geophys. Res.*, **92**, 4467-4490.
Heyn, M.F., Biernat, H.K., Rijnbeek, R.P. & Semenov, V.S. (1988) *J. Plasma Physics*, **40**, 235-252.
Hones, E.W., Jr. (1984) In *Magnetic Reconnection in Space and Laboratory Plasmas*, Geophys. Monogr. Ser. vol. 30, (edited by E.W. Hones, Jr.), Washington D.C. : AGU, pgs 178-184.
Kennel, C.F. (1987) Invited paper for the International 'Rosenbluth Symposium' on Dynamics of Particles and Plasmas, Austin, Texas, February 5-6, 1987.
Kennel, C.F., Coroniti, F.V., Scarf, F.L., Livesey, W.A., Russell, C.T., Smith, E.J., Wenzel, K.P. & Scholer, M. (1986) *J. Geophys. Res.*, **91**, 11917-11928.
Lockwood, M. & Cowley, S.W.H. (1988) *Adv. Space Res.*, **8** (9), 281-299.
Lockwood, M., Sandholt, P.E. & Cowley, S.W.H. (1989) *Geophys. Res. Lett.*, **16**, 33-36.
Lockwood, M., Cowley, S.W.H. & Freeman, M.P. (1990a) *J. Geophys. Res.*, **95**, 7961-7972.
Lockwood, M., Cowley, S.W.H. & Sandholt, P.E. (1990b) *Eos*, **71**, 709.
Paschmann, G. (1984) In *Magnetic Reconnection in Space and Laboratory Plasmas*, Geophys. Monogr. Ser. vol. 30, (edited by E.W. Jones, Jr.), Washington D.C. : AGU, pgs. 114-123.
Paschmann, G., Papamastoakis, I., Baumjohann, W., Sckopke, N. Carlson, C.W., Sonnerup, B.U.Ø. & Luhr, H. (1986) *J. Geophys.Res.*, **91**, 11099-11115.
Petschek, H.E. (1964) In *AAS NASA Symposium on the Physics of Solar Flares*, **NASA SP-50**, 425-439.
Pilipenko, V.A. (1990) *J.Atmos.Terr.Phys.*, in press.
Potemra, T.A. & Saflekos, N.A. (1979) In *Magnetospheric Boundary Layers*, (edited by B. Battrick), Noordwijk-Holland: ESA SP-148, pgs. 193-198.
Priest, E.R. & Forbes, T.G. (1986) *J. Geophys. Res.* **91**,5579-5588.
Priest, E.R. & Lee, L. (1990) *J Plasma Phys.* **44**, 337-360.
Pu Song, Elphic, R.C. & Russell, C.T. (1988) *Geophys. Res. Lett.*, **15**, 744-747.
Pu Song, Russell, C.T., Gosling, J.T., Thomsen, M.& Elphic, R.C. (1990) *J. Geophys. Res. Lett.*, in press.
Rich, F.J. & Maynard, N.C. (1989) *J.Geophys. Res.*, **94**, 3687-3701.

Richard, R.L., Walker, R.J., Sydora, R.D. & Ashour-Abdalla, M. (1989) *J. Geophys. Res.*, **94**, 2471-2483.
Rijnbeek, R.P., Cowley, S.W.H., Southwood, D.J. & Russell, C.T. (1984) *J. Geophys. Res.*, **89**, 786-800.
Rijnbeek, R.P., Biernat, H.K., Heyn, M.F., Semenov, V.S., Farrugia, C.J., Southwood, D.J., Paschmann, G., Sckopke, N. & Russell, C.T. (1989) *Annales Geophysicae*, **7**, (3), 297-310.
Roelof, E.C. (1987) *Geophys. Res. Lett.*, **14**, 652-655.
Roelof, E.C., Mitchell, D.G. & Williams, D.J. (1985) *J.Geophys. Res.*, **90**, 10991-11008.
Rostoker, G. & Eastman, T. (1987) *J.Geophys. Res.*, **92**, 12187-12201.
Russell, C.T. & Elphic, R.C. (1978) *Space Sci. Rev.*, **22**, 681-715.
Sandholt, P.E., Lockwood, M., Oguti, T., Cowley, S.W.H., Freeman, K.S.C., Lybekk, B., Egeland, A. & Willis, D.M. (1990) *J. Geophys. Res.*, **95**, 1039-1060.
Sato, T. & Iijima, T. (1979) *Space Sci. Rev.*, **24**, 347-366.
Saunders, M.A. (1983) *J.Geophys.*, **52**, 190-198.
Saunders, M.A. (1989a) *Geophys. Res. Lett.*, **16**, 151-154.
Saunders, M.A. (1989b) *Antarctic Science*, **1**, 193-203.
Scholer, M. (1988) *Geophys. Res. Lett.*, **15**, 291-294.
Schwartz, S.J. (1985) In *Solar System Magnetic Fields*, (edited by E.R.Priest), Dordrecht: D.Reidel. pgs.190-223.
Siscoe, G.L. (1983) In *Solar Terrestrial Physics - Principles and Theoretical Foundations*, (edited by R.L.Carovillano and J.M.Forbes), Dordrecht-Holland: D. Reidel. pgs 11-100.
Shi, Y. & Lee, L.C. (1990) *Planet. Space Sci.*, **38**, 437-458.
Smith, R.A., Goertz, C.K. & Grossmann, W. (1986) *Geophys. Res. Lett.*, **13**, 1380-1383.
Sonnerup, B.U.O., Paschmann, G., Papamastorakis, I., Sckopke, N., Haerendel, G., Bame, S.J., Asbridge, J.R., Gosling, J.T. & Russell, C.T. (1981) *J. Geophys. Res.*, **86**, 10049-10067.
Southwood, D.J. & Hughes, W.J. (1983) *Space Sci. Rev.*, **35**, 301-366.
Southwood, D.J., Farrugia, C.J. & Saunders, M.A. (1988) *Planet.Space Sci.*, **36**, 503-508.
Stone, R.G. & Tsurutani, B.T. (1985) Editors of *Collisionless Shocks in the Heliosphere: A Tutorial Review*, Geophys.Monogr.Ser. vol.34, Washington, D.C.: AGU.
Terasawa, T. & Scholer, M. (1989) *Science*, **224**, 1050-1057.
Tsurutani, B.T. & Stone, R.G. (1985) Editors of *Collisionless Shocks in the Heliosphere: Reviews of Current Research*, Geophys.Monogr.Ser. vol. 35, Washington, D.C.: AGU.
Tsyganenko, N.A. (1987) *Planet.Space Sci.*, **35**, 1347-1358.
Vasyliunas, V.M. (1975) In *The Magnetospheres of the Earth and Jupiter*, (edited by V.Formisano), Dordrecht-Holland: D.Reidel. pgs.179-185.
Walker, R.J. & Ogino, T. (1989) *ISEE Transactions on Plasma Science*, **17**, 135-149.
Wenzel, K.-P., Domingo, V. & Schmidt, R. (1987) *ESA Bulletin*, No.50, 8-16.
Williams, D.J. (1990) In *Proceedings of the International Conference on Auroral Physics* held in Cambridge, UK, July 1988.

17

MHD Forces in Astrophysical Disks and Jets

J HEYVAERTS

Université Paris, 92195 Meudon, France

Abstract

Accretion disks and jets are found to be associated with several kinds of astrophysical objects which are briefly described. Observational evidence and theoretical arguments indicating that accretion disks are likely to be magnetized are presented, and the consequences for the structure and evolution of disks are discussed. It is shown that dynamo action in the disk results in a magnetized coronal halo around it. Being stressed by the disk's differential rotation, the halo is heated by a process similar to solar coronal heating. Conversely, Lorentz forces are exerted on the disk, which cause non-local interaction and redistribution of angular momentum. The disk, even at relatively low magnetization, may become non-keplerian, and eventually differential rotation may come to be frozen. A model incorporating these effects is outlined.

If the disk is threaded by an open field configuration, the latter can support a rotating magnetized wind. It is shown that such a wind can remove angular momentum from the disk, thus allowing a secular contraction of it to take place. Finally, the shape of the wind in regions remote from its source is discussed. Recent results (Heyvaerts and Norman, 1989) show that the asymptotic form of the momentum equation perpendicular to flux surfaces implies that, for almost all such winds, the asymptotic structure consists of a family of nested cylindrical surfaces, potentially explaining the collimation of the observed "jets", in particular the famous jets from active galactic nuclei, without resorting to any external confinement.

17.1 The Link Between Disks and Jets

Observations in the past 20 years have provided evidence for a variety of objects accreting from their environment, which may consist of diffuse matter or some mass-losing companion. This is certainly so for X-ray binaries and cataclysmic variables (Frank *et al*, 1985), for young stellar objects (Rodriguez 1988, Adams *et al*, 1987) and is quite possible in the case of Active Galactic Nuclei, AGN's (Malkan, 1983; Shields, 1977, 1978).

A number of these objects, though not all, show jet-type outflows. This is well documented for young stellar objects (see Strom *et al*, 1988 or Königl, 1989) and for AGN's (see Begelman *et al*, 1984). The evidence is dimmer in the case of X-ray binaries or cataclysmic variables (Mauche and Raymond, 1987). SS433, which has a spectacular relativistic jet (Margon 1984) is associated with a binary system, probably with an accretion disk (Leibowitz *et al* 1984).

The centre of our galaxy itself harbours an interesting "molecular torus" orbiting around the dynamical centre (Genzel, 1989). An X-ray emitting gas (Watson *et al*, 1981), present in diffuse form should be the driver of a wind from the galactic centre, as yet undetected.

The disk/jet association cannot be fortuitous. Some colleagues believe that the jet is just a wind emanating from the accreting object, which should then be anisotropic, and forced by its interaction with the disk into directions perpendicular to the disk (see Ferrari *et al*, 1985), thus providing some collimation. Others think that the disk (or the wind-emitting object) could be threaded by a large-scale, open, magnetic field, so that some magnetized, rotating, wind would be emitted.

It is often suggested that such winds should be centrifugally driven, tapping directly the orbital energy of the accreting flow. This, in our opinion, need not be so, since the same energy can be responsible for strong plasma heating, forming a hot corona, which might be thermally driven into a wind, just as well. In both cases, accretion has the potential of explaining the source of the energy of the wind. Rotating winds also carry to infinity, in material and magnetic form, a significant amount of angular momentum (Pudritz and Norman, 1986; Königl, 1989). The classical view of accretion disks includes a viscous transport of angular momentum outwards to the edges of the disk, where it accumulates, spreading out, or returns to a reservoir, like the companion star with which it interacts tidally (King, 1989). The removal of angular momentum by a wind offers an alternative process, and for this reason might turn out to be an essential element of the accretion machine, not just a secondary effect. Magnetic stresses in the wind can also cause very efficient collimation. We come back to this point in some detail below.

17.2 Structure and Origin of the Disk's Magnetic Field

The concept of a magnetized wind implies that some large-scale field spans the disk, at least partly. There is evidence for strong, ordered, fields at most of these astrophysical sites. Neutron stars are known to have a very large magnetic field, which might span, by turbulent transport, at least the innermost edge of accretion disks. In fact the existence of such magnetospheres may prevent the formation of an accretion disk in short-period cataclysmic variables (Hameury *et al* 1988, Lasota *et al* 1986). In star-forming regions, the magnetic field has been measured by Zeeman splitting of the 18 cm OH line (Troland *et al* 1986) and shown to be of order 10-100 microgauss. Its ordered character is apparent from starlight polarization observations, and its direction is almost aligned with the axes of jet outflows observed in this region (Vrba *et al* 1988, Strom *et al* 1988). The polarization of the optical light emitted (by synchrotron emission) from the jets of M87 and 3C66B also suggests a rather ordered magnetic field with some internal structure (Fraix Burnet *et al* 1990 a-b). In our own galactic centre, the innermost features observed in the infra-red and radio range are likely to be threaded by an organized magnetic field of rather high value (17.39 milligauss) (Aitken, 1989). Values much higher than a few microgauss, typical of the interstellar medium, are also deduced, some 10pc away from the centre, from radio polarization observations (Sofue *et al* 1987). The MHD situation in this region of our galaxy, a quite interesting topic, is reviewed by Heyvaerts (1988).

The presence of strong fields in accretion disks could *a priori* be regarded as a natural consequence of a concentration of magnetized matter, since the magnetic Reynolds numbers involved are high, and one could naively think that the flux freezing theorem should be applicable. Closer examination shows things to be more complicated.

Of course, if the material is only partly ionized, account should be taken of ambipolar diffusion (Nakano, 1978; Königl, 1989). This leak of neutral matter through the field-tied ionized component is weak for densities less than $10^{11} cm^{-3}$. Above $10^{11} cm^{-3}$ various mi-

crophysics phenomena increase the field diffusion greatly (Nakano, 1988; Norman and Heyvaerts, 1985).

Of much more singificance, however, is the fact that accretion disks should be turbulent, since their viscous Reynolds number, based on molecular viscosity is so high. As a result, there appears not only an effective turbulent viscosity, ν_*, but also a turbulent magnetic diffusion, η_*, and more generally all the effects which appear in the context of mean-field electrodynamics (Moffatt, 1978). For isotropic turbulence, ν_* and η_* are both almost equal to the product $v_t \ell_t$ of the eddy velocity v_t and the correlation length ℓ_t. In the standard Shakura Sunyaev (1973) hypothesis, ℓ_t is the disk thickness and v_t is a fraction of the sound speed. This leads to a characteristic decay time (for those components of the field that vary on the disk thickness scale) of the order of the keplerian period. Those which vary on the radial scale decay in the accretion time-scale. In both cases, the relevant "effective" magnetic Reynolds number is of the order of unity, and turbulent transport plays an important role.

Since helicity effects on the field generation, the so-called α-effect, self-consistently accompany the effective diffusivity, an $\alpha-\omega$ dynamo could be expected in the differentially rotating disk. An analysis of the linear $\alpha-\omega$ dynamo regime in a simple model of accretion disks has been presented long ago by Pudritz (1981 a-b), and, as far as we are aware, the theory has not evolved very much since then, apart from the recognition of the role of buoyancy (Stella and Rosner, 1984).

The work of Pudritz assumes subsonic turbulence and makes use of the so-called first-order smoothing approximation; the boundary conditions at the border of the dynamo-active region (the interior of the disk) are assumed to be those of a potential field. The evolution equation for the large-scale field is then:

$$\frac{\partial B}{\partial t} = \mathrm{rot}(\mathbf{v} \times \mathbf{B} + \alpha^* \mathbf{B}) + \eta^* \nabla^2 \mathbf{B}$$

where the turbulent transport coefficients α^*, η^*, and the large-scale velocity field $\mathbf{v}$ are supposedly known functions of space. Pudritz estimated them, on the assumption that $v_t \approx M_t c_s$ and $\ell_t \approx h$, where M_t is the turbulence Mach number. The possibility of obtaining self-generated fields depends on the ratio of the field-generating terms to dissipation, and is summarized in the requirement that the dynamo

number, defined as:

$$Q = \left(\frac{h^2 \Omega}{\eta^*}\right)\left(\frac{h\alpha^*}{\eta^*}\right) \qquad (17.1)$$

exceed some critical value, which is found by looking for non- trivial, time-independent, solutions of the dynamo equation. Pudritz found that the critical value of Q is a few units, while Q itself is almost equal to $(17.50\pi)M_t^{-2}$. Thus, there is dynamo action if turbulence is "sufficiently subsonic". Since it is not known how subsonic exactly it is in reality, it is impossible to draw any quantitative conclusion. The same is true of the dynamo modes, the B_z field being then given by a Bessel function with a radial scale that varies between the disk thickness and large values (near threshold, especially).

This work demonstrates the plausibility of dynamo action. The (major) difficulties concerned with nonlinear effects, such as the reaction of the dynamo field on the flow or on α^*, realistic radial and vertical dependences of α^* and η^*, and the role of buoyancy are no easier here than they are in the stellar context, and await considerable progress in this field (Weiss, 1989). Since our aim is not to review this difficult subject, we leave it here.

17.3 Disk Coronae

Were it of dynamo origin, or just convected in the disk by partial field freezing, a magnetic field may well be present in actual accretion disks. If so, it will also thread some region sandwiching the disk. Stresses are exerted on this outer region by small-scale perturbations created by turbulence in the disk, and by large-scale shearing, much in the same way as the solar corona is acted upon by the small- and large-scale motions of the photosphere. These stresses should dissipate by processes similar to those studied in the solar context (for recent reviews see Priest, 1990), and be an efficient source of heating, thus creating a disk's corona as first pointed out by Galeev *et al* (1979). The role and the X-ray emissivity of the corona have been largely studied since then, for example by Kuperus and Ionson (1985), Burm (1986) or Zuccarello *et al* (1987). These authors take the view that the disk remains essentially keplerian, and that the thickness of the corona is of the order of the disk thickness.

It is our feeling that the corona could, in reality, be a lot thicker, even though "primary loops", which emerge by buoyancy out of the disk (Stella and Rosner, 1984), are of this characteristic size. This

is because reconnection between small loops gives larger loops, and so on. So, statistically, large loops could form from small ones. This is just the well-known inverse cascading process which, as we also know, is likely to exist in 3D MHD, when turbulence is endowed with kinetic or magnetic helicity. Then, the corona is likely to have long-range magnetic connections between different parts of the disk.

If magnetization is large enough, it is also likely that the effect of Lorentz forces on the disk dynamics will become non-negligible. This point has been recognized long ago, and, assuming that such stresses are primarily exerted inside the disk by small-scale features which permanently undergo stretching and reconnection, it was shown by Coroniti (1981) that the net effect could be incorporated in some "effective viscosity". Belvedere and Molteni (1984) suggest that the magnetic tension inside the disk also introduces a deviation from the keplerian law.

The magnetic connection of widely separated regions of the disk by field lines passing through the corona and closing on the disk is the potential cause of non-local interaction forces between different parts of the disk, as well as the source of stresses in the corona, and therefore of its heating. Priest and I considered this situation, as described below.

17.4 Magnetic Coupling between a Disk and its Corona

Heyvaerts and Priest (1989) took simultaneous account of the heating which results from the turbulent dissipation of stresses exerted by the disk on the corona due to relative shearing of the footpoints of field lines and of the back reaction of these stresses on the dynamics of the disk. In full rigour this involves the simultaneous solution of three problems: (a) The disk dynamics, including large-scale forces from the corona; (b) The flux transport problem in the disk, most probably with turbulent dynamo action; (c) The heating and magnetic structure of the coronal region.

Of course some simplifications were necessary. We did not solve problem (b) self-consistently, but only assumed a distribution of vertical component $B_z(r)$. We paid some attention to the compatibility of our chosen model with the equations it was supposed to satisfy. Problem (c) is the most formidable of the three. We took inspiration from earlier results on the decline of turbulence (Taylor, 1986), which we earlier used to develop a calculation of coronal heating (Heyvaerts and Priest, 1984). Taylor's results indicate that magnetic turbulence

endowed with magnetic helicity ($\int \mathbf{A}\cdot\mathbf{B}dV$) decays to a linear force-free field. Of course a system like the disk's corona that is permanently stressed never reaches this totally relaxed limit. Nevertheless, if the characteristic turbulence decay time is short enough, as compared to the winding up time (the keplerian period) we may expect the magnetic structure not to differ very much from this state. We then assumed the coronal structure to be a linear force- free field, and took the associated α as a parameter of the model, since its value is physically controlled by a gain/loss balance which appeared to be very difficult to assess properly. In a stationary state, the rate of heating of the corona will be the Poynting flux, which can be calculated once the dynamics of the disk are solved. Assuming a thin disk, with surface density $\sigma(r)$, angular velocity $\Omega(r)$ and radial velocity $v(r)$, these dynamics are described by height-integrated equations for the conservation of mass, momentum and energy, incorporating the magnetic stresses. These have been written in detail by Lovelace *et al* (1987). Assuming a field with B_r, B_ϕ odd in z and even B_z, the equations of conservation of mass, radial and vertical momentum, angular momentum and energy, respectively, take the form

$$2\pi r\sigma v = -\dot{M} \tag{17.2}$$

$$\sigma\Omega^2 r + 2/\mu B_r^+ B_z = \sigma GM/r^2 \tag{17.3}$$

$$p(r) = (B_r^{+2} + B_\phi^{+2})/2\mu \tag{17.4}$$

$$\frac{1}{r}\frac{d}{dr}r(\sigma r^2 v\Omega - \sigma v_* r^2\frac{d\Omega}{dr}) = r\frac{2B_\phi^+ B_z}{\mu} \tag{17.5}$$

$$\frac{1}{r}\frac{d}{dr}r(\frac{1}{2}\sigma r^2\Omega^2 v + \sigma\phi_G v + v_*\sigma r^2\Omega\frac{d\Omega}{dr}) = 2r\Omega\frac{b_z B_\phi^+}{\mu} - 2\sigma_B T^4. \tag{17.6}$$

In these expressions, B_z^+ and B_ϕ^+ are the values of B_z and B_ϕ as $z \to 0$ through positive values. B_z and B_ϕ are discontinuous when crossing the disk when the zero thickness limit is taken. As a result, a finite tension force is exerted, similar to the one which holds solar prominences in equilibrium. In the present context this force, which is a novel feature of the disk/corona interaction, has two components one of which appears in (17.3) and is responsible for deviations to the Kepler law. The other one appears in (17.5) as an external torque, which allows angular momentum exchanges between the disk and the corona. Since the field is force-free in the corona, rB_ϕ is constant on a field line, and $2\pi rB_z$ has opposite values on magnetically conjugate regions on the disk. From this it follows that there is no net angular momentum transport to the closed corona, since the vertical

angular momentum fluxes at two magnetically conjugate locations are opposite to each other. So what is "sent" by one is "received" by the other, and these stresses only redistribute angular momentum between conjugate points. In the energy equation we recognise on the left-hand side the divergence of fluxes of kinetic, gravitational and viscous energies. On the right-hand side we find the radiative losses per unit surface and the Poynting flux, which measures the energy exchanged with the corona in electro-mechanical form. The energy sent by the disk to the corona per unit disk surface is then, at distance r from the centre:

$$f_H = -2r\Omega \frac{B_z B_\phi^+}{\mu} \tag{17.7}$$

Note that in the non-magnetized limit equation (17.3) implies keplerian rotation, while the elimination of v^* between (17.5) and (17.6), neglecting angular momentum flux towards the acrreting object, gives the familiar result

$$2\sigma_B T^4 = \frac{3}{4\pi}\frac{GM\dot{M}}{r^3} \tag{17.8}$$

In the magnetized case, equation (4.4) can be similarly integrated to give, neglecting angular momentum flux to the central object,

$$\dot{M}r^2\Omega + 2\pi\nu^*\sigma r^3\frac{d\Omega}{dr} + \int_{R*}^{r} 2\pi x dx \frac{2\pi B_z B_\phi^+}{\mu} = 0 \tag{17.9}$$

Now, the problem can be completely solved: we assume a normal component of the field at the disk, $B_z(r)$; then the linear force- free field assumption, which is meant to be a sketchy representation of the state of the stressed corona, gives the field in the corona, in particular $B_r^+(r)$ and $B_\phi^+(r)$. These, then, are known functions of r in eqn (17.3) and (17.9). Once these are solved for $\Omega(r)$ and $\sigma(r)$, mass conservation will give the radial velocity and energy equation will give the radiative losses of the disk, while the integration over the disk of the Poynting flux to the corona will give the rate of coronal heating. It is possible to show that the latter necessarily is a positive quantity if $\Omega(r)$ is a decreasing function.

Let us then concentrate on the radial motion and angular momentum equation. It is useful to normalize variables. Let R be the disk radius, and B_0 a representative value of B_z, say. Let

$$x = r/R; \qquad \mathbf{b} = \mathbf{B}/B_0; \qquad \omega^2 = \Omega^2/(GM/R^3) \tag{17.10}$$

The viscosity ν^* is also a dimensional quantity. Note that if the disk were non-magnetized, equation (17.9) would give the surface density

as

$$\sigma_0 = \dot{M}/3\pi\nu^* \tag{17.11}$$

and, for a constant ν^*, the mass of the disk would be:

$$M_{DO} = \dot{M}R^2/3\nu^*. \tag{17.12}$$

It is then pictorial to use, as a typical value of ν^*, $R^2\dot{M}/3M_D$, where M_D is a mass representative of the disk mass. Then, let

$$\nu = 3M_D\nu^*/R^2\dot{M} \qquad \text{and} \qquad \Sigma = \sigma/\sigma_0. \tag{17.13}$$

Inserting these dimensionless variables into the equations causes two dimensionless parameters to appear, namely

$$\lambda = (4\pi R^3 B_0^2/\mu)/(\frac{GMM_D}{R}) \tag{17.14}$$

$$\mu = (4\pi R^3 B_0^2/\mu)\sqrt{\frac{GM}{R^3}}/(\frac{GM\dot{M}}{R}) = \lambda/(\dot{M}/M_D\sqrt{\frac{GM}{R^3}}) \tag{17.15}$$

Here λ is the ratio of the disk magnetic energy to its potential gravitational energy. The parameter μ compares two rates of dissipation: one, the numerator, is a gross estimate of the rate of coronal heating (all the magnetic energy burnt in one twist period), the other, the denominator, is of the order of the gravitational energy lost in passing a substantial fraction of the disk radius. Note that μ equals λ multplied by the ratio of the accretion time-scale ($M_D/\dot{M}$) to the keplerian time-scale. This ratio is a large number. The equation for σ and ω are:

$$\Sigma = -\frac{\lambda}{2}\frac{b_z b_{r+}}{x(\omega^2 - 1/x^3)} \tag{17.16}$$

$$\frac{d\omega}{dx} = \frac{(\omega^2 - 1/x^3)(\omega x^2 + \mu\int_0^x \omega^2 b_z(u)b_\phi^+(u)du)}{\frac{\lambda}{3}x^2 b_r^+(x)b_z(x)} \tag{17.17}$$

Since Σ and $(d\omega/dx)$ should not be singular, it is necessary that the numerators and denominators vanish together. The product $(b_r^+ b_z)$ vanishes at arcade boundaries ($b_r^+ = 0$) or at polarity reversals ($b_z = 0$). At these points, for Σ to remain finite, $\omega = x^{-3/2}$, the keplerian value. A careful study of critical points reveals that the acceptable solution for $\omega(x)$ is the critical solution at $b_z = 0$, $\omega = x^{-3/2}$. For small magnetization it gives a solution as shown in Fig. 17.1a, slower than keplerian at the innermost of two conjugate points, and faster at the other. If the magnetization is large enough that μ reaches a critical value (which depends as α^{-1} on the force-free field parameters α) a horizontal tangent appears at one of the polarity reversals. At this point the magnetic stresses "freeze" the

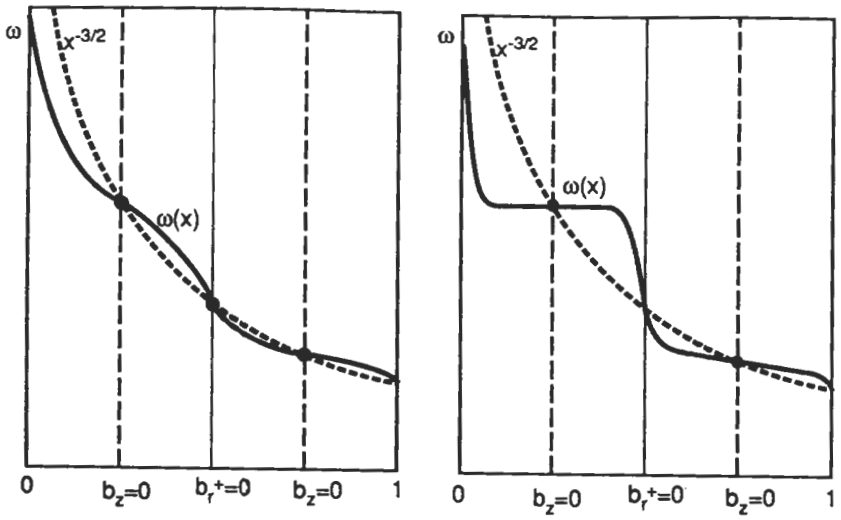


Figure 17.1. Angular velocity (ω) was a function of radial distance (x) for (a) small and (b) critical magnetization.

differential rotation into solid body rotation. It is not possible to pursue the analysis consistently with the same model beyond this point.

The rate of coronal heating is given by

$$H \approx \frac{4\pi B_0^2 R^3}{\mu}\sqrt{\frac{GM}{R^3}}(R\alpha)\chi \qquad (17.18)$$

where α is the force-free parameter and χ is a number of order unity which depends on the details of the magnetic configuration.

This shows that, when the limit is approached where the magnetic tensions are able to freeze the differential rotation, the coronal heating becomes equal to the disk emission, in order of magnitude, if $R\alpha \approx 1$.

17.5 Basic Theory of the MHD Forces Acting on Jets

As explained before, winds are expected to be associated with disks if the latter are threaded by open magnetic fields. These winds may be driven centrifugally (Blandford and Payne, 1982) or else could be driven thermally if a corona forms around the accretion disk, as we have described in the previous two sections.

It is quite striking, observationally, just how collimated the jets from active galactic nuclei are along their axes. Is it possible that this beaming is the result of MHD forces?

One interesting possibility would be that, in the asymptotic regions of the wind flow, the pinching force associated with the azimuthal component of the field ($B_\theta^2/\mu_0 r$), the so-called hoop stress, become large enough to balance gas and magnetic pressure gradients, thus providing collimation.

It is important to note that the question of the shape of the flow in axisymmetric, rotating, MHD winds is a difficult subject, which, until recently, has been avoided by assuming that the flux surfaces have known shapes, for example that they are nested cones (Weber and Davis 1967). Since then the problem of calculating from first principles the very shape of flux surfaces has been approached numerically (Sakurai, 1985, 1987) and analytically, in simplified situations (Blandford and Payne, 1982).

Very recently (Heyvaerts and Norman, 1989), some general analytical results have been obtained, to which the last part of this review is dedicated.

Let us first remind the reader of the basics of this problem. A stationary axisymmetric flow is sought. A cylindrical coordinate system is used, where the z-axis is the axis of rotation. All vectorial quantities can be split into poloidal (in a meridional plane) and toroidal (or azimuthal) components. Since $\partial/\partial\theta = 0$, the equations $div\mathbf{B} = 0$ and $div(\rho\mathbf{v}) = 0$ imply that

$$div\mathbf{B}_p = 0 \qquad \text{and} \qquad div(\rho\mathbf{v}_p) = 0. \qquad (17.19)$$

This allows us to write

$$\mathbf{B}_p = -\frac{1}{r}\frac{\partial a}{\partial z}\,\hat{\mathbf{r}} + \frac{1}{r}\frac{\partial a}{\partial r}\,\hat{\mathbf{z}}, \qquad (17.20)$$

where the surface $a(r,z) =$ constant is a flux surface. The vector $\rho\mathbf{v}_p$ is given by a similar expression involving another function of r and z (the flow function). It is physically obvious that in a stationary state, the flow surfaces must coincide with the flux surfaces if flux

freezing holds true. Actually, the

$$\mathrm{rot}(\mathbf{v}\times\mathbf{B})=\mathbf{O} \quad \text{implies} \quad \mathbf{v}\times\mathbf{B}=+\nabla\phi. \tag{17.21}$$

Then $\mathbf{v}_p\times\mathbf{B}_p=0$, and since both $\rho\mathbf{v}_p$ and $\mathbf{B}_p$ have zero divergence, we find:

$$\rho\mathbf{v}_p=\alpha(a)\mathbf{B}_p. \tag{17.22}$$

The poloidal components of $\mathbf{v}\times\mathbf{B}=\nabla\phi$ lead after simple manipulation to:

$$(v_\theta-\frac{\alpha(a)B_\theta}{\rho})\frac{1}{r}\nabla a=\nabla\phi \tag{17.23}$$

from which it follows that $\nabla\phi$ is parallel to ∇a, and so ϕ is a function of a. Let $\Omega=d\phi/da$. The above equation then gives

$$\rho v_\theta=\alpha(a)B_\theta+\rho r\Omega(a). \tag{17.24}$$

Altogether, the fluid motion consists of a solid body rotation of each magnetic surface (at its own rotation frequency) on which a field-aligned flow is superposed.

$$\mathbf{v}=r\Omega(a)\hat{\theta}+\alpha(a)\frac{\mathbf{B}}{\rho}. \tag{17.25}$$

The equation of motion can similarly be split into a toroidal and a poloidal part. The former can easily be brought to the form:

$$\nabla a\times\nabla(r\alpha(a)v_\theta-r\frac{B_\theta}{\mu})=0 \tag{17.26}$$

which has the consequence that

$$rv_\theta-r\frac{B_\theta}{\mu\alpha(a)}=L(a) \tag{17.27}$$

where $L(a)$ is a surface function which has the dimensions of specific angular momentum (per unit mass), but it does not reduce only to "matter" angular momentum. It has also a magnetic part, which represents the effect of torques exerted on the matter by $(\mathbf{j}\times\mathbf{B})$ forces.

The poloidal part of the equation of motion, still a vectorial equation, can be projected along the direction of $\mathbf{B}_p$ and orthogonal to it. This gives, on the one hand, another first integral, the specific energy, which is again a surface function (the Bernoulli equation)

$$E(a)=\frac{1}{2}v_p^2+\frac{1}{2}v_\theta^2+\frac{\gamma}{\gamma-1}\frac{p}{\rho}+G-\frac{\Omega(a)rB_\theta}{\mu\alpha(a)} \tag{17.28}$$

where here G is the gravitational potential. Some tedious calculations brings the component perpendicular to $\mathbf{B}_p$ to the form (the

so-called transfield equation):

$$\begin{aligned}&\frac{\alpha}{\rho r}\left[\frac{\partial}{\partial z}\frac{\alpha}{\rho r}\frac{\partial a}{\partial z}+\frac{\partial}{\partial r}\frac{\alpha}{\rho r}\frac{\partial a}{\partial r}\right]-\frac{1}{\mu\rho}\frac{1}{r}\left[\frac{\partial}{\partial z}\frac{1}{r}\frac{\partial a}{\partial z}+\frac{\partial}{\partial r}\frac{1}{r}\frac{\partial a}{\partial r}\right]\\&=E'(a)-\frac{Q'(a)}{\gamma-1}\rho^{\gamma-1}+\frac{\alpha'}{\alpha}\frac{(L-r^2\Omega)}{(\mu\alpha^2-\rho)^2}\frac{\mu\alpha^2\rho}{r^2}\\&-\frac{\rho}{r^2}\frac{(L'-r^2\Omega')(L-r^2\Omega)}{\mu\alpha^2-\rho}-\frac{LL'}{r^2}.\end{aligned} \tag{17.29}$$

The isorotation law and the angular momentum conservation give v_θ and B_θ in terms of poloidal variables:

$$v_\theta=\frac{L}{r}+\frac{\rho}{r}\frac{L-r^2\Omega}{\mu\alpha^2-\rho} \tag{17.30}$$

$$B_\theta=\mu\alpha\frac{\rho}{r}\frac{L-r^2\Omega}{\mu\alpha^2-\rho}. \tag{17.32}$$

This shows that a regular solution must have:

$$\rho=\rho_A=\mu\alpha^2 \tag{17.33}$$

when

$$r^2=r_A^2=L/\Omega. \tag{17.34}$$

Finally, the requirement that the thermodynamics be adiabatic gives

$$\frac{p}{\rho^\gamma}=Q(a). \tag{17.35}$$

Note that the first integrals, $\Omega(a)$, $Q(a)$, $L(a)$, $\alpha(a)$, $E(a)$ are fixed not only by the boundary conditions at the source of the wind, but also by the requirement that the solution goes smoothly through the critical points.

In the limit of no flow the constants $\alpha(a)$ and $\Omega(a)$ approach zero, while $L(a)$ tends to infinity, with αL approaching $(-rB_\theta/\mu)$. The transfield equation then reduces to the well-known Grad-Shafranoff equation of magnetohydrostatics.

If the shape of flux surfaces, that is the function $a(r,z)$ which the transfield equation is supposed to yield, were known, the Bernoulli equation would just reduce to an algebraic equation for the poloidal flow, since v_θ and B_θ can be expressed in terms of ρ, the pressure, p, is given by the adiabatic law, and v_p^2 is related to other functions by:

$$v_p^2=\frac{\alpha^2}{\rho^2}B_p^2=\frac{\alpha^2}{\rho^2}\frac{\nabla^2 a}{r^2}. \tag{17.36}$$

Actually, when these substitutions are made, the equation (17.28)

takes the form:

$$\frac{1}{2}\frac{\alpha^2}{\rho^2 r^2}|\nabla a|^2 = E(a) - G(r,z) - \frac{\gamma}{\gamma-1}Q(a)\rho^{\gamma-1} - \frac{1}{2}\frac{L^2}{r^2} + 2\frac{L\Omega}{\rho_A-\rho}\rho + \frac{L\Omega^2\rho}{(\rho_A-\rho)^2} - \frac{L^2}{r^2}\frac{\rho}{\rho_A-\rho} - \frac{1}{2}\frac{L^2}{r^2}\left(\frac{\rho}{\rho_A-\rho}\right)^2 - \Omega^2\frac{r^2\rho}{\rho_A-\rho} - \frac{1}{2}\frac{\Omega^2 r^2\rho^2}{(\rho_A-\rho)^2} \tag{17.37}$$

which, for known $|\nabla a|$, relates, for a given a, and for a given set of first integrals, the density ρ to the position, r. As is well known, the topology of the solutions is such that consistent solutions (large density at small r, vanishing pressure at large r, regularity and continuity of the solution) can be found only if the solution goes through 2 critical points (fast and slow). This imposes 2 conditions on the first integrals $(Q, \Omega, L, \alpha, E)$. All solutions $\rho(r)$ are such that $\rho = \rho_A$ at $r = r_A$, so this does not introduce any extra condition by itself.

Ideally, one would like to be able to find the solution of the Bernoulli equation for ρ in terms of $|\nabla a|$, substitute in the transfield and solve for $a(r,z)$, but this of course is impossible in practice.

Moreover, solving the transfield equation for $a(r,z)$ by itself also introduces extra constraints of criticality type. This can best be seen by performing the following manipulations. Since the transfield equation describes the shape of the fieldlines, while the Bernoulli equation gives the modulus of $\mathbf{v}_p$, which is related to that of $\mathbf{B}_p$, it seems interesting to write:

$$\mathbf{B}_p = B_p\hat{\mathbf{t}}, \qquad \mathbf{v}_p = v_p\hat{\mathbf{t}} \tag{17.38}$$

where $\hat{\mathbf{t}}$ is the unit vector tangent to poloidal field lines $(\cos\phi, 0, \sin\phi)$, and to convert the transfield equation into an equation for the direction field $\hat{\mathbf{t}}(r,z)$. This is relatively easy using the identity:

$$\frac{\alpha}{\rho r}\left\{\frac{\partial}{\partial z}\frac{\alpha}{\rho r}\frac{\partial a}{\partial z} + \frac{\partial}{\partial r}\frac{\alpha}{\rho r}\frac{\partial a}{\partial r}\right\} = -\frac{\nabla a}{|\nabla a|^2}\left((v_p\cdot\nabla)v_p - \nabla\frac{v_p^2}{2}\right), \tag{17.39}$$

$$\frac{1}{r}\left\{\frac{\partial}{\partial z}\frac{1}{r}\frac{\partial a}{\partial z} + \frac{\partial}{\partial r}\frac{1}{r}\frac{\partial a}{\partial r}\right\} = -\frac{\nabla a}{|\nabla a|^2}\left((B_p\cdot\nabla)B_p) - \nabla\frac{B_p^2}{2}\right). \tag{17.40}$$

Using the relation which defines curvature of a planar curve

$$(\mathbf{t}\cdot\nabla)\mathbf{t} = \mathbf{n}\frac{d\phi}{ds}$$

and

$$-\mathbf{n}|\nabla a| = \nabla a$$

the transfield equation (17.29) may be rewritten as an equation for

the curvature of the poloidal field line:

$$\frac{1}{|\nabla a|}\left(\left(v_p^2 - \frac{B_p^2}{\mu\rho}\right)\frac{d\phi}{ds} - \left((n\cdot\nabla)\frac{v_p^2}{2} - \frac{1}{\mu\rho}(n\cdot\nabla)\frac{B_p^2}{2}\right)\right) = E'(a) - \frac{Q'(a)}{\gamma-1}\rho^{\gamma-1} + \frac{\alpha'}{\alpha}\frac{(L-r^2\Omega)^2}{(\mu\alpha^2-\rho)^2}\frac{\mu\alpha^2\rho}{r^2} - \frac{\rho}{r^2}\frac{(L'-r^2\Omega')(L-r^2\Omega)}{(\mu\alpha^2-\rho)} - \frac{LL'}{r^2}. \tag{17.41}$$

Here, v_p^2, or equivalently B_p^2, can be obtained from the Bernoulli equation, and we have

$$v_p^2 - \frac{B_p^2}{\mu\rho} = v_p^2\left(1 - \frac{B_p^2}{\mu\rho v_p^2}\right) = v_p^2\left(1 - \frac{\rho}{\mu\alpha^2}\right). \tag{17.42}$$

Then the transfield equation can be written as

$$\frac{v_p^2}{|\nabla a|}\left(1 - \frac{\rho}{\mu\alpha^2}\right)\frac{d\phi}{ds} = \text{Something Complicated.} \tag{17.43}$$

This shows that we have to impose an extra regularity condition, which is that the complicated right-hand side of the above equation should vanish when $\rho = \mu\alpha^2$, that is also when $r^2 = L/\Omega$. This is the "Alfvén regularity condition". If it is not satisfied, the flux surfaces would have a kink at the Alfvèn point (infinite curvature). Some tedious algebra reduces the vanishing of the right-hand side of the above equation of $r = r_A$ to the "simple" form:

$$\frac{d\ log\ \alpha}{da} + 2(1-p)\frac{d\ log\ r_A}{da} - 2(1-p)\frac{sin\ \phi_A}{r_A|\nabla a|_A} + \frac{E(a)}{v_{pA}^2}\frac{d\ logE}{da} - \frac{1}{(\gamma-1)}\frac{Q\rho_A^{\gamma-1}}{v_{pA}^2}\frac{d\ log\ Q}{da} + \frac{\Omega^2 r_A^2}{v_{pA}^2}\left(\frac{1}{(1-p)^2}\frac{d\ log\ \alpha}{da} + \frac{2p}{1-p}\frac{d\ log\ r_A}{da} - \frac{d\ log\ \Omega}{da}\right) = 0 \tag{17.44}$$

where p depends on a and is defined by:

$$p(a) = 1 + \frac{1}{2}\left[\left(\partial\ log\ \rho/\partial\ log\ r\right)_a\right]_{r=r_A}. \tag{17.45}$$

Note that here the partial derivative is taken at fixed a, and its value is calculated ar $r_A(a)$. This Alfvén regularity condition links the solution of the Bernoulli equation, on line a, which determines the run of $\rho(r)$ at fixed a (i.e. $p(a)$ in particular) and the cross-field variations of the various first integrals.

17.6 Focusing of MHD Winds

Fortunately, it is not necessary to solve the problem completely to obtain results concerning the general shape of flux surfaces.

Look, for example at the Bernoulli equation in the following form

$$\frac{1}{2}v_p^2 = E(a) - G(r,z) - \frac{\gamma}{\gamma-1}Q(a)\rho^{\gamma-1} - \frac{1}{2}\frac{L^2}{r^2} + 2\frac{L\Omega}{\rho_A-\rho}\rho + \frac{L\Omega\rho^2}{(\rho_A-\rho)^2} - \frac{L^2}{r^2}\frac{\rho}{\rho_A-\rho} - \frac{1}{2}\frac{L^2}{r^2}\left(\frac{\rho}{\rho_A-\rho}\right)^2 - \Omega^2\frac{r^2\rho}{\rho_A-\rho} - \frac{1}{2}\frac{\Omega^2r^2\rho^2}{(\rho_A-\rho)^2}. \tag{17.46}$$

Note that if r is to approach infinity on the field line considered, ρ, being bounded by ρ_A, can actually only approach zero, in such a way that ρr^2 remains finite, since otherwise the penultimate term of the right-hand side would diverge negatively. Then, as $r \to \infty$, we can deduce that the limit of v_p^2 at given a, $v_\infty(a)$, is (with self-explanatory notation!)

$$\lim_{\infty/a}\left(\frac{1}{2}v_p^2\right) = E - \Omega^2\lim_{\infty/a}\left(\frac{\rho r^2}{\rho_A}\right). \tag{17.47}$$

Let us define $R^2(r,z)$ and $R_\infty^2(a)$ by:

$$\rho r^2 = \rho_A R^2(r,z) \qquad \text{and} \qquad \lim_{\infty/a} R^2 = R_\infty^2(a). \tag{17.48}$$

This limit may vanish; we consider that later on. However, we can be sure that, for each a, there exists an upper bound to ρr^2, since $\rho_A R_\infty^2$ cannot exceed $E(a)$. Then

$$\rho r^2 \le K(a). \tag{17.49}$$

This can be converted into an inequality for $|\nabla a|$, since

$$\rho^2 v_p^2 = \alpha^2 B_p^2 = \alpha^2\frac{\nabla a^2}{r^2}. \tag{17.50}$$

Then we find the inequality

$$|\nabla a| \le \frac{\lambda(a)}{r} \tag{17.51}$$

where $\lambda(a)$ is some finite function. From this it is easy to deduce a constraint on the shape of the field lines, since at a given r, the distance between two such lines, a_1 and a_2, is

$$|z(a_2) - z(a_1)| = \left|\int_{a_1}^{a_2} dz\right| = \int_{a_1}^{a_2}\frac{da}{\left|\frac{\partial a}{\partial z}\right|} \ge r\int_{a_1}^{a_2}\frac{da}{\lambda(a)}. \tag{17.52}$$

This shows that the relative divergence of flux surfaces cannot be

less than a linear function of r. Flux surfaces such that $\lim(z/r) = 0$ are impossible. We say that the asymptotics cannot be flat, that is (z/r) approaches a finite or an infinite value. We show that it indeed necessarily approaches an infinite value.

To prove that, let us consider now the transfield equation, which really contains the information on these shapes. Suppose we consider a set of field lines for which $r \to \infty$ (it can be on asymptotically conical or parabolic branches). In this limit, we reduce the equation which gives the curvature of poloidal field lines (eqn 17.41) for $r >> r_A$, to

$$r\frac{d\phi}{ds} = \frac{\mu}{2\alpha v_p}\frac{\partial}{\partial a}(\alpha^2\Omega^2R^4) - r\cos\phi\frac{\Omega^2}{v_p^2}\frac{\partial R^2}{\partial z} \tag{17.53}$$

where the function $R(a,z)$ is defined by (17.48).

It can be checked that as $z \to \infty$ the second term on the right-hand side approaches zero, since we note that:

$$r\cos\phi\frac{\partial R^2}{\partial z} = \frac{r}{z}\cos\phi\frac{\partial R^2}{\partial \ell nz}. \tag{17.54}$$

Since (r/z) and $\cos\phi$ are at best finite, and since $\partial R^2/\partial \ell nz$ approaches zero because R^2 must approach a finite value, we see that the transfield equation asymptotically approaches the form:

$$r\frac{d\phi}{ds} = \frac{\mu}{2\alpha v_p}\frac{\partial}{\partial a}(\alpha^2\Omega^2R_\infty^4). \tag{17.55}$$

Now, the right-hand side is a function of a, and since the derivative with respect to the curvilinear coordinate is at constant a, the asymptotic shape of the flux surface labelled a is given by the system of equations:

$$\begin{cases} r\frac{d\phi}{ds} = \psi(a), \\ dr = cos\phi ds \end{cases}. \tag{17.56}$$

This system can be integrated analytically. It has no physically admissible solution, unless $\psi(a) = 0$. Therefore we conclude that the transfield equation boils down, in the asymptotic limit, to the single "solvability condition".

$$\frac{d}{da}(\alpha^2\Omega^2R_\infty^4) = 0. \tag{17.57}$$

We therefore arrive at the rather interesting conclusion, that the quantity $\alpha\Omega R_\infty^2$ is the same on all field lines that go to $r = \infty$:

$$\lim_{\infty/a}\alpha\Omega R_\infty^2 = C \tag{17.58}$$

where C is a constant, independent of a.

This has an important physical implication. Let us recall that

$$rB_\theta = \mu\alpha\frac{L - r^2\Omega}{\mu_0\alpha^2 - \rho}\rho. \tag{17.59}$$

As $r \to \infty$, this reduces for $r >> r_A$, $\rho << \rho_A$ to:

$$\lim_{\infty/a} rB_\theta = -\mu_0\alpha\Omega \lim \frac{r^2\rho}{\rho_A} = -\mu_0\alpha\Omega R_\infty^2. \tag{17.60}$$

This shows that the current enclosed in a flux surface a approaches a value independent of a as $r \to \infty$ on conical or paraboloidal flux surfaces. Or else, that the poloidal current approaches zero on those flux surfaces which approach $r \to \infty$.

As a rule the total distributed poloidal current in a hemisphere is non-zero. This can be checked for examples, and it is given by the limiting value of $2\pi r B_\theta$ as a, the flux surface index, approaches the equatorial value, A, by values slightly poleward of it. Evaluating this in the limit $r << r_A$, we obtain:

$$I_p \quad \text{hemisphere} \approx -2\pi\mu \lim_{(a\to A-0)} \alpha(a)L(a). \tag{17.61}$$

For a symmetric wind and an antisymmetric polarity in the two hemispheres, $\alpha(A+\epsilon) = -\alpha(A-\epsilon)$, so that a current sheet forms in the equator, which closes the total current circuit from infinity. This current sheet is a well-known feature of the solar wind. In general, however, the poloidal current in one hemisphere, equator excluded, as defined by (17.61) does not vanish since $\alpha(A) \neq 0$ if there is wind at the equator and $L(a) = \Omega(a)$ is in general non-zero too, there. In the case of a wind source of infinite extent, like an infinite disk, it is less obvious that αL should differ from zero at the outermost field line, which is at infinity.

From this we conclude that the asymptotic shape of a wind that carries a non-vanishing distributed poloidal current in each hemisphere cannot exclusively consist of conical or parabolic flux surfaces, since such surfaces have been shown to carry no poloidal current to infinity. Then such flows necessarily contain in their central regions a set of nested cylindrical flux surfaces which carry the poloidal current.

That a wind cannot simply asymptote to nested cones can also be seen rather directly, since, if we denote by $\phi_\infty(a)$ the angle defined by $\lim_{\infty/a}(z/r) = tg\phi_\infty(a)$, we can show easily that for this geometry:

$$r|\nabla a| \to -\cos\phi_\infty(da/d\phi_\infty) \tag{17.62}$$

and since

$$\lim(r|\nabla a|) = (\mu\alpha v_\infty R_\infty^2)^{-1} \tag{17.63}$$

the asymptotic direction $\phi_\infty(a)$ are given by the differential equation

$$\frac{1}{\cos\phi_\infty}\frac{d\phi_\infty}{da} = -\frac{1}{\mu_0\alpha(a)v_\infty(a)R_\infty^2(a)} \tag{17.64}$$

which is to be coupled to the solvability condition (17.58)

$$\mu_0\alpha\Omega R_\infty^2 = C \tag{17.65}$$

and the constant is different from zero since on some of these lines, at least, $R_\infty \neq 0$. Substituting for R_∞^2 from (17.65) in the angle/flux equation, we obtain

$$\frac{1}{\cos\phi_\infty}\frac{d\phi_\infty}{da} = -\frac{\Omega(a)}{Cv_\infty(a)} \tag{17.66}$$

which integrates, polewards from equator, to give

$$\tan\phi_\infty(a) = \sinh\left(\int_A^a \frac{\Omega(a')}{Cv_\infty(a')}da'\right). \tag{17.67}$$

Unless $v_\infty = 0$ at the polar field line, and provided the wind source object spans a finite flux ($A \neq \infty$), the integral cannot diverge, and there is no way to build a global solution which fills all space at infinity with conical flux surfaces only, as expected.

Also, it is possible to show that a conical zone cannot smoothly join a parabolic zone. This can be seen by first proving that if $R_\infty(a)$ is a strictly positive function on the interval $[a_1, A]$, bounds included, then the corresponding field lines are asymptotically confined, for a in this interval, between two cones. This comes from the fact that the strict positivity of $R_\infty(a)$ and its finiteness put upper and lower limits to $r|\nabla a|$

$$\mu(a) \leq r|\nabla a| \leq \lambda(a) \tag{17.68}$$

where $\lambda(a)$ and $\mu(a)$ are also strictly positive and finite, and this in turn implies that the level line $a(r,z) = a$ is asymptotically confined between cones of opening angle

$$tg\bar{\phi}(a) = \sinh\left(\int_a^A \frac{da}{\mu(a)}\right) \tag{17.69}$$

$$tg\underline{\phi}(a) = \sinh\left(\int_a^A \frac{da}{\lambda(a)}\right). \tag{17.70}$$

Since the total flux A in the wind is finite by assumption, and $\mu(a)$ strictly positive, these slopes $tg\underline{\phi}$ and $tg\bar{\phi}$ are finite. In brief, where $R_\infty(a)$ remains strictly positive from equator on, $\phi_\infty(a)$ cannot reach $(\pi/2)$ to make a smooth connection to a parabolic zone. To make such a connection R_∞ should approach zero at the upper boundary of the conical zone, thus allowing $\mu(a) \to 0$ at this boundary and

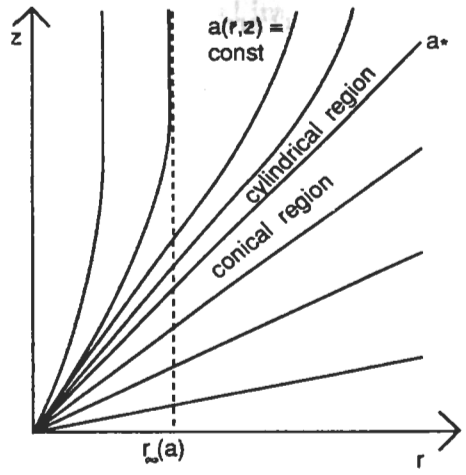


Figure 17.2. Streamlines or fieldlines that are a mixture of cones and cylinders

$\bar{\phi}(a) \to \pi/2$. But in the conical zone, the constant C of eqn (17.65) is non-zero and forbids R_∞ to approach zero unless if it happens that $\alpha \to \infty$ or $\Omega \to \infty$, which are both unphysical. One could think of R_∞ being discontinuous, but this would involve an unphysical jump of poloidal current (current sheet) at infinity.

So, cones cannot co-exist with paraboloids.

The above discussion has also shown that, for a source object spanned by a finite flux, parabolic asymptotics implies that the constant C which appears in the limit transfield equation (17.65) be zero, since it cannot reach zero from non-zero values at the equator, and (17.65) itself shows that, if $R_\infty = 0$ somewhere in the asymptotic, non-cylindrical, zone, it must be zero at all the other places in this same zone, since $\alpha(a)$ and $\Omega(a)$ are likely to vanish nowhere.

We next ask whether such winds could asymptote to a mixture of nested cones and cylinders. Such a blend of asymptotic behaviour, surprisingly is perfectly conceivable from the mathematical point of view, and it is relatively easy to construct explicit examples, similar to those of Figure 17.2, which represents a series of level lines of some hypothetical $a(r,z)$. This of course demands that when the transition value between the cylindrical and conical region is approached from the cylindrical side, we have

$$\lim_{(a \to a_* - 0)} (r_\infty(a)) = \infty. \tag{17.71}$$

Therefore, at least along the field lines close to that limit, and for very large z's, it should be true that the geometry is cylindrical,

while $r_\infty(a) >> r_A(a)$. In this limit the transfield equation can be written as (we replace r_∞ by r)

$$\frac{\alpha}{\rho r}\frac{d}{dr}\frac{\alpha}{\rho r}\frac{da}{dr} = E'(a) + \frac{\alpha'}{\alpha}\frac{r^2\rho\Omega^2}{\mu\,\alpha^2} - \frac{r^2\rho\Omega\Omega'}{\mu\,\alpha^2} \tag{17.72}$$

while, in the same limit $r >> r_A$, the Bernoulli equation is:

$$\frac{1}{2}\frac{\alpha^2}{\rho^2 r^2}\left(\frac{da}{dr}\right)^2 = E(a) - \frac{r^2\rho\Omega^2}{\mu\alpha^2}. \tag{17.73}$$

This is a system for $\rho(r)$ and $a(r)$ which can be brought to quadratures, giving for v_z (that is for $v_\infty(r)$):

$$v_\infty^2 = 2E(a) - \frac{k\Omega(a)}{\alpha(a)} \tag{17.74}$$

where k is an integration constant. It is, by the way, no surprise to find that this constant is just, using (17.73) and (17.74)

$$\frac{k}{2} = \frac{r^2\rho\Omega}{\mu\,\alpha} = \alpha\Omega R_\infty^2, \tag{17.75}$$

which means that cylindrical flux surfaces that reach $r >> r_A$ approximately share the properties of those for which $r \to \infty$. The function $a(r)$ itself turns out to be given by the quadrature:

$$\frac{dr}{r} = \frac{2}{k\mu}\frac{\Omega da}{\sqrt{2E - k\frac{\Omega}{\alpha}}}. \tag{17.76}$$

To obtain a consistent solution, for which $r \to \infty$ at some value, a_*, of a, it is necessary that the integral

$$\int_0^{a_*} \frac{\Omega(a')da'}{v_\infty(a')}$$

diverges as $a \to a_*$, which means that the constant k takes a value such that $(2E - k\Omega/\alpha)$ has a double zero. This means that v_∞ should vanish at the boundary between the two zones. So by continuity, it should vanish also when this boundary is approached from the conical side. But this is impossible for flows which exceed the fast-mode speed, because the latter must exceed $B_\theta^2/(\mu_0\rho)$

$$c_f^2 \geq B_\theta^2/(\mu_0\rho). \tag{17.77}$$

However, we have shown that on conical field lines

$$rB_\theta \to -\mu_0\alpha\Omega R_\infty^2 = -\mu_0 C. \tag{17.78}$$

Then, in such regions

$$c_f^2 \geq \mu_0\frac{C^2}{\lim(\rho r^2)} = \frac{\mu_0\alpha^2\Omega^2R_\infty^2}{\mu_0\alpha^2R_\infty^2} = \Omega^2R_\infty^2 \tag{17.79}$$

and, since C^2 is strictly positive and ρr^2 bounded, c_f^2 cannot approach

zero in any such region of the flow, and so also not v_∞, which should be larger than c_f.

The remaining possibility is that all field lines be asymptotically cylindrical if the total poloidal current in one hemisphere is non-vanishing.

One could conceive a solution where the wind is confined, that is, where no flux surface at all reaches infinity, the flow being pinched by a B_θ field, self-created by the poloidal current of the jet, so that no flow would reach to $r > r_\infty(A)$, the limit radius of the equatorial field line (a singular line). However, if $v_\infty \neq 0$ on this field line, this would mean also $B_p \neq 0$ there, and then a jump of B_z at the jet boundary, which means again a (possibly unphysical) current sheet. It is perhaps a matter of debate whether indeed such a current sheet, which would wrap around the boundary of the flow should be regarded as more unphysical than the equatorial current sheet (the existence of which is certain) and to which it would connect. We propose however another solution below, in our view more satisfactory.

Let us seek for a solution in which $r_\infty(a)$ would approach infinity as $a \to A$, the equatorial value. First provisionally ignore the transfield equation, and assume such a geometry of flux surfaces is given. Then note that the consequence is that v_∞ should indeed approach zero on the equatorial field lines, since, as a rule, we can write the poloidal velocity as:

$$v_p = \frac{\alpha}{\rho} B_p = \frac{\alpha}{\rho r^2} r|\nabla a| \to \frac{\alpha}{\lim(\rho r^2)} (r_\infty \frac{da}{dr_\infty}). \qquad (17.80)$$

For conical asymptotics $r|\nabla a|$ could not approach zero, but in the assumed cylindrical geometry, with $r_\infty(A) \to \infty$, the quantity $(da/d\log r_\infty)$ certainly approaches zero. ρr^2 in the denominator remains finite, because we can write for $r >> r_A$ (applicable to field lines for which $r_\infty >> r_A$)

$$\lim(rB_\theta) = -\lim \frac{\rho r^2 \Omega}{\alpha} \qquad (17.81)$$

and then on equatorial flux surface

$$\rho r^2 \to (I_{Pol}(A)\alpha(A))/(2\pi\Omega(A)) \qquad (17.82)$$

which is finite. Using this in (17.80), we see that $v_\infty(A) = 0$ for such a geometry.

It remains to show that the envisaged geometry is indeed consistent with the transfield equation. For $r >> r_A$, the latter is expressed by (17.76) in this geometry. We note that $E(a), \Omega(a)$ and $\alpha(a)$ should be even functions near $a = A$, by symmetry between the two hemispheres, $(E\alpha/\Omega)$ should have an extremum there, and a constant k exists which endows the function $(2E - k\Omega/\alpha)$ with a double zero there. We checked that near the axis, where the approximation $r >> r_A$ breaks down, the solution (17.76) with the "divergent" value of k could be relayed by another approximate solution which allows it to connect to $r = 0$ for $a = 0$ (on the axis). This makes it likely that indeed a consistent solution could be found for all a's in $[0, A]$.

A final remark is in order. We note that, on the equatorial field line, as $r \to \infty$, the fast-mode speed obeys the following inequality:

$$c_f^2 \geq B_\theta^2/\mu_0\rho = \frac{I_P^2}{4\pi^2\mu_0}(\rho r^2)^{-1}.$$

Since (ρr^2) approaches a finite value, c_f^2 is itself finite. But we have just shown that in this geometry $v_\infty \to 0$ on the equatorial flux surface. Therefore, it could certainly not be super-fast mode there. This indicates that in the considered geometry, the fast critical point is rejected to infinity as the equatorial flux surface is approached. The latter then appears as rather singular.

In summary, our work (Heyvaerts and Norman, 1989) shows the necessity for a magnetized wind, carrying a non-vanishing poloidal current in one hemisphere, to converge at infinity to a structure of nested cylindrical flux surfaces.

We believe this result to give strong support to the idea that jets of active galaxies indeed are magnetically collimated rotating magnetized winds. The extension of our results to relativistic winds is in progress (Chiueh, Begelman and Li, 1990).

References

Adam, F C, Lada, C J, Shu, F H, 1987, *Astrophys. J.* **312**, 788.

Aitken, D K, 1989, in "The Center of our Galaxy", IAU Symp. 136, (Mark Morris ed.), p.457.

Begelman, M C, Blandford, R D and Rees, M J, 1984, *Rev. Mod. Phys.* **56**, 265.

Belvedere, G and Molteni, D, 1984, *Il Nuovo Cimento* **7C**, 17.54.

Blandford, R D and Payne, D G, 1982, *Mon. Not. Roy. Astr. Soc.* **199**, 883.

Burm, H, 1986, *Astron. Astrophys.* **165**, 120.

Chiueh, T, Li, Z Y and Begelman, M C, 1990, preprint.

Coroniti, F V, 1981, *Astrophys. J.* **244**, 17.25.

Ferrari, A, Trussonni, E, Rosner, R, Tsinganos, K, 1985, *Astrophys. J.* **294**, 397.

Fraix Burnet, D, Le Borgne, J F, Nieto, J L, 1990, *Astron. Astrophys.* , in press.
Fraix Burnet, Nieto, J L, Poulain, P, 1990, *Astron. Astrophys. Letts.*, in press.
Franck, J, King, A R, Raine, D J, 1987, *Accretive power in astrophysics*, Cambridge University Press.
Galeev, A A, Rosner, R, Vaiana, G S, 1979, *Astrophys. J.* **229**, 318.
Genzel, R, 1989, in "The Center of our Galaxy", *loc. cit.*, p.393.
Hameury, J M, King, A R, Lasota, J P, 1988, *Astron. Astrophys.* **195**, L12.
Heyvaerts, J, 1989, in "The Center of our Galaxy", *loc. cit.*, p.301.
Heyvaerts, J and Norman, C A, 1989, *Astrophys. J.* **347**, 1055.
Heyvaerts, J and Priest, E R, 1984, *Astron. Astrophys.* **137**, 63.
Heyvaerts, J and Priest, E R, 1989 *Astron. Astrophys.* **216**, 230.
King, A R, 1989, in "Accretion disks and magnetic fields in astrophysics" hereafter "Noto Conference", G Belvedere, ed., Kluwer Academic Press, p.43.
Königl, A, 1989, in "Noto Conference" *loc. cit.*, p.165.
Kuperus, M and Ionson, J A, 1985, *Astron. Astrophys.* **148**, 309.
Lasota, J P, Hameury, J M, King, A R, 1989, in "Noto Conference" *loc. cit.*, p.107.
Leibovitz, E M, Mazeh, T, Mendelsohn, H, 1984, *Nature* **307**,341.
Lovelace, P V E, Wang, J C L, Sulkanen, M E, 1987, *Astrophys. J.* **315**, 17.22.
Malkan, M A, 1983, *Astrophys. J.* **268**, 17.20.
Margon, B, 1984, *Ann. Rev. Astron. Astrophys.* **2**, 17.25.
Mauche, C W and Raymond, J C, 1987, *Astrophys. J.* **325**, 690.
Moffat, H K, 1978, "Magnetic field generation in electrically conducting fluids", Cambridge University Press.
Nakano, T, 1979, *Pub. Astron. Soc. Jap.*, **31**, 17.52.
Nakano, T, 1988, in "Galactic and Extragalactic Star Forming Regions", R Pudritz and M Finch, ed., Dordrecht, p.111.
Norman, C A and Heyvaerts, J, 1985, *Astron. Astrophys.* **147**, 247.
Priest, E R, 1990, Proc. IAU Symp. **142** on *Basic Plasma Processes on the Sun*, (ed. E.R. Priest and V. Krishan), p271.
Pudritz, R E, 1981a, *Mon. Not. Roy. Astron. Soc.* **195**, 881.
Pudritz, R E, 1981b, *Mon. Not. Roy. Astron. Soc.* **195**, 897.
Pudritz, R E and Norman, C A, 1983, *Astrophys. J.* **274**, 17.52.
Pudritz, R E and Norman, C A, 1986, *Astrophys. J.* **301**, 17.19.
Rodriguez, L F, 1988, in "Galactic and Extragalactic Star Formation" *loc. cit.*, p.97.
Sakurai, T, 1985, *Astron. Astrophys.* **152**, 121.
Sakurai, T, 1987, *Pub. Astron. Soc. Jap.*, **39**, 821.
Shakura, N I and Sunyaev, R A, 1973, *Astron. Astrophys.* **24**, 337.
Shields, G A, 1977, *Astrophys. Letters* **18**, 119.
Shields, G A, 1978, *Nature*, **272**, 706.
Sofue, Y, Reich, W, Inoue, M, Seiradakis, J H, 1987, *Pub. Astron. Soc. Jap.* **39**, 95.
Stella, R, Rosner, R, 1984, *Astrophys. J.* **277**, 312.
Strom, S E, Strom, K M and Edwards, S, 1988, in "Galactic and Extragalactic Star Formation Regions" *log. cit.*, p.53.
Taylor, J B, 1986, *Rev. Mod. Phys.* **58**, 741.
Troland, T H, Crutcher, R M and Kazés, I, 1986, *Astrophys. J.* **304**, L57.
Van Ballegooijen, A A, 1989, in "Noto Conference" *loc. cit.*, p.99.
Vrba, F J, Strom, S E and Strom, K M, 1988, *Astron. J.*, **96**, 680.

Watson, M G, Willingale, R, Grindlay, J E, Hertz, P, 1981, *Astrophys. J.* **250**, 142.
Weber, E J and Davis, L, Jr, 1967, *Astrophys. J.* **148**, 217.
Weiss, N O, 1989, in "Noto Conference" *loc. cit.*, p.11.
Zuccarello, F, Burm, H, Kuperus, M, Raadu, M, Spicer, D S, 1987, *Astron. Astrophys.* **180**, 218.